# Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 24 (2022), Issue 9    ISSN 1438-8871    Editor in Chief: Gunther Eysenbach, MD, MPH

## Contents

### Reviews

## Original Papers

## Viewpoints

## Corrigenda and Addendas

## Research Letter

Review

# Use of Social Media Data to Diagnose and Monitor Psychotic Disorders: Systematic Review

Alban Lejeune[1], MD; Benoit-Marie Robaglia[2], MS; Michel Walter[1,3], MD, PhD; Sofian Berrouiguet[1,4], MD; Christophe Lemey[1,3,5], MD

[1]Unité de Recherche Clinique Intersectorielle, Hôpital de Bohars, Centre Hospitalier Régional Universitaire de Brest, Bohars, France

[2]Institut Polytechnique de Paris, Palaiseau, France

[3]Faculté de Médecine et Sciences de la Santé, Université de Bretagne Occidentale, Brest, France

[4]Laboratoire de Traitement de l'Information Médicale, Unité Mixte de Recherche 1101, Institut National de la Santé et de la Recherche Médicale, Brest, France

[5]Lab-STICC, Unité Mixte de Recherche, Centre National de la Recherche Scientifique 6285, F-29238, École Nationale Supérieure Mines-Télécom Atlantique Bretagne Pays de la Loire, Brest, France

**Corresponding Author:**
Alban Lejeune, MD
Unité de Recherche Clinique Intersectorielle
Hôpital de Bohars
Centre Hospitalier Régional Universitaire de Brest
Route de Ploudalmézeau
Bohars, 29820
France
Phone: 33 6389910008
Email: alban.lejeune@gmail.com

## Abstract

**Background:** Schizophrenia is a disease associated with high burden, and improvement in care is necessary. Artificial intelligence (AI) has been used to diagnose several medical conditions as well as psychiatric disorders. However, this technology requires large amounts of data to be efficient. Social media data could be used to improve diagnostic capabilities.

**Objective:** The objective of our study is to analyze the current capabilities of AI to use social media data as a diagnostic tool for psychotic disorders.

**Methods:** A systematic review of the literature was conducted using several databases (PubMed, Embase, Cochrane, PsycInfo, and IEEE Xplore) using relevant keywords to search for articles published as of November 12, 2021. We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria to identify, select, and critically assess the quality of the relevant studies while minimizing bias. We critically analyzed the methodology of the studies to detect any bias and presented the results.

**Results:** Among the 93 studies identified, 7 studies were included for analyses. The included studies presented encouraging results. Social media data could be used in several ways to care for patients with schizophrenia, including the monitoring of patients after the first episode of psychosis. We identified several limitations in the included studies, mainly lack of access to clinical diagnostic data, small sample size, and heterogeneity in study quality. We recommend using state-of-the-art natural language processing neural networks, called language models, to model social media activity. Combined with the synthetic minority oversampling technique, language models can tackle the imbalanced data set limitation, which is a necessary constraint to train unbiased classifiers. Furthermore, language models can be easily adapted to the classification task with a procedure called "fine-tuning."

**Conclusions:** The use of social media data for the diagnosis of psychotic disorders is promising. However, most of the included studies had significant biases; we therefore could not draw conclusions about accuracy in clinical situations. Future studies need to use more accurate methodologies to obtain unbiased results.

***(J Med Internet Res 2022;24(9):e36986)*** doi:10.2196/36986

**KEYWORDS**

schizophrenia; psychotic disorders; psychiatric disorders; artificial intelligence; AI; machine learning; neural network; social media

## *Introduction*

### Background

Schizophrenia is a chronic mental disease affecting 20 million people worldwide [1]. Although treatment with medicine and psychosocial support is effective, people with schizophrenia are less likely to seek treatment. According to the World Health Organization, efforts to transfer care from mental health institutions to the community need to be accelerated [2]. Currently, schizophrenia is a disease associated with high burden [3,4], and efforts should be taken to reduce this burden.

Artificial intelligence (AI) has emerged as a way to improve several medical tasks [5,6]. AI algorithms can identify patterns in a data set to generate diagnostic tools. In other medical disciplines, AI has already shown good accuracy for diagnostic purposes. It can match current diagnostic capabilities in some specific fields. In psychiatry, AI could be used for diagnostic purposes to support daily patient assessment or drug prescription [7]. AI has also been studied to improve diagnostic and classification capabilities [8]. Additionally, it has been used in suicide risk detection [9] and mood disorder diagnoses [10,11].

Despite encouraging results, it is still unclear how AI should be implemented in clinical practice. The potential of this technology is not yet fully understood. AI could be used to bring completely new tools into health care. We believe that social media might be used to create new diagnostic or monitoring tools. Indeed, AI requires a large database to extract a patient's profile [12], and social media platforms provide very broad sources of information. People can disclose personal information on social media, including health-related information. Studies have used these data to identify broad human traits (such as intelligence or personality traits) [13]. Subtle features of everyday language could be analyzed to predict mental diseases [14-16]. Prior works showed that social media data can be used for risk classification associated with mental health diseases, such as suicide risk [17]. AI can be used to detect posts associated with mental illness [18]. Therefore, we chose to study the use of AI applied to social media because we believe it could become a brand-new tool in the care of patients with psychotic disorders.

### Focus on AI Technologies

AI can be used to perform several different tasks. Machine learning algorithms are generally classified as supervised or unsupervised learning. The main type of machine learning algorithms used in the included studies was supervised learning. Supervised learning algorithms allow patterns correlated to a result to be determined in a data set [12]. The supervised algorithms are separated into two categories: regression and classification. Classification algorithms allow data to be classified into separate categories. Patterns can be used to classify patients in a given group. Decision tree, support vector machine (SVM), and random forest can perform classification tasks. Regression algorithms are used to predict quantitative data. Logistic regression and LASSO (least absolute shrinkage and selection operator) regression are part of this class [6].

Artificial neural networks (NNs) are powerful AI tools built in reference to the cortical neural structure. They can perform supervised or unsupervised tasks. NNs are organized in a succession of layers, with each layer having its input on the output of the previous one. Information travels from the input neurons to the hidden layers before arriving at the outcome layer where the final decision is made [5]. In an NN, each layer functions differently. This type of AI requires significant computing power and large databases [19].

The most used machine learning algorithm in our study is the SVM [20]. The idea of this algorithm is to learn a linear separation (a hyperplane) of the data points to classify them. As there are infinite hyperplanes satisfying this condition, the SVM algorithm learns a hyperplane with the maximum margin, the maximum distance between the classes. However, as most real-life data sets cannot be linearly separated, the SVM uses what is called the "kernel trick." This transformation projects the initial data points in a higher dimensional feature space where the new is linearly separable. However, there are many limitations to consider when using the SVM: (1) finding a good kernel function is difficult in practice, (2) training is time-consuming on large data sets, and (3) the model is very sensitive to the hyperparameters.

Overfitting means fitting an AI model on data noise or error instead of the actual relationship. It represents one of the limitations of AI. Overfitting is either due to having a small data sample or too many parameters compared to the data [12]. Cross-validation is one of the techniques used to reduce overfitting. With this technique, the data set is split into several groups that are themselves divided into training data and validation data. Therefore, for each group, the statistical model is trained and then validated by a different data set. This technique reduces the risk of having an overoptimistic estimate [21]. Other techniques such as the dropout rate are also used to reduce overfitting. Dropout is a regularization technique for NNs to reduce overfitting and improve generalization [22]. The idea is to randomly ignore neurons (and their connections) from the NN during training. Thus, as the NN architecture is changing at every inference, the same input data can produce a different output. The intuition is that it forces the units to be less codependent and more robust. The main difference between cross-validation and dropout lies in the source of randomness; in cross-validation, the data are randomly split into training and validation sets whereas in dropout, the neural units are randomly discarded.

We have explained the machine learning parameters used in this paper in Multimedia Appendix 1 [23-25].

### Prior Work and Goal of This Study

Currently, the diagnosis of psychotic disorders can be subjected to delay. These delays can vary depending on where the patient lives. When the patient remains untreated with psychotic symptoms, there can be important social consequences, including a risk of violence in some cases [26]. The duration of untreated psychosis could have a significant impact on the patient's psychosocial condition. Early detection and treatment could help improve the care of patients with psychotic disorders [27]. Later, during the evolution of the disease, being able to diagnose a relapse sooner could have significant impact on the patient's quality of life and reduce caregiver burden.

AI has been studied for the diagnosis of several psychiatric conditions, including schizophrenia. A Korean team used machine learning to identify patterns on CT (computed tomography) scans and classify patients [28]. One of the most robust studies on the classification of psychotic disorders using machine learning comes from an American team [8] that used a clustering algorithm to build 3 biotypes using clinical data and laboratory measures. Neuroimaging and social functioning measures were used for validation. Beyond the main initial goal, this study showed that several new sources of data can be used to improve diagnostic capabilities. In our study, we examined social media data as a new source of data. The symptoms of schizophrenia are very broad, and some of these symptoms could impact patients' social media activities. For example, we can hypothesize that delusion and disorganized speech or behavior could be seen on social media posts. Alterations in language are being increasingly studied in schizophrenia [15,29] and could be used for the detection of psychotic disorders on social media.

Several studies have shown the capabilities of AI in identifying mental health diseases on social media, with most of them published in the last few years. One brief report [30] reviewed the literature available until December 2020. We hypothesize that social media data could be used to follow patients with schizophrenia or patients at risk of psychosis to identify the first psychotic episode or a relapse of psychotic disorders sooner. Patients could have access to care before psychotic symptoms overtake their social functioning capabilities. Thus, we could reduce the burden of schizophrenia.

### Objective of This Review

The objective of our study is to analyze prior works on the use of social media data with machine learning to diagnose a psychotic episode. The diagnostic capabilities would be studied in a broad sense, including the diagnosis of relapse. Therefore, we performed a systematic review of the literature and critically evaluated the included articles and their methodology.

## Methods

### Search Strategy

We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria to identify, select, and critically assess relevant studies while minimizing bias. We searched the bibliographic databases PubMed, Embase, Cochrane, PsycInfo, and IEEE Xplore for studies published until November 12, 2021. We based the keywords list on 3 fields: schizophrenia, AI, and social networks. The search strategy is described in Multimedia Appendix 2. To limit the selection bias, we did not apply any restriction in terms of population. Papers found by any other means were included if they met the inclusion criteria. Studies that were not published in English were excluded.

### Study Selection

We included clinical trials and observational studies. The primary objective was to include studies using AI to identify users with a psychotic disorder on social media. Given the low number of published studies on this subject, studies related to any psychotic disorder were included. Studies were selected by 2 independent authors. We excluded studies using social media posts as control data to study language alterations in schizophrenia.

### Data Collection Process

Data were extracted from each paper independently using a standard form by 1 reviewer. The following information was collected: the main author's name and country of origin, year of publication, population, social media and technology used, features, inclusion and exclusion criteria, main objective, method, main endpoint, results, and main limitations.

### Synthesis Method

The results of the selected studies will be presented as graphs and tables. Machine learning studies often use different parameters. The graphs will group studies using similar parameters. No secondary analysis of statistical data will be performed. The results presented will comprise only those presented by the authors of the included studies. The results will be presented with the parameters used in the articles. The machine learning parameters used in this paper are explained in Multimedia Appendix 1.

### Risk of Bias and Quality Assessment

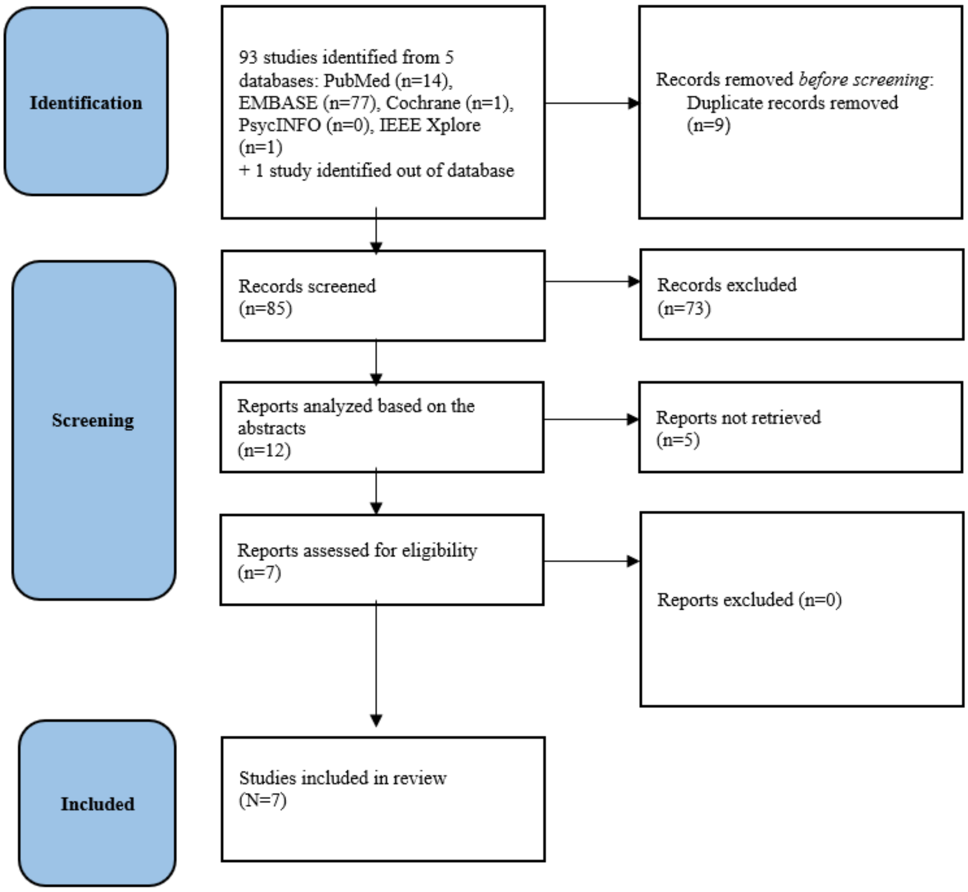
The quality of the included studies will be assessed using the PRISMA certainty tool. This research will be conducted using the PRISMA checklist (Multimedia Appendix 3). The machine learning methodology of the articles will be evaluated by an author experienced in AI. Risk of bias will be critically assessed by all the authors.

## Results

### Flowchart

We developed a PRISMA flowchart summarizing the steps of the review (Figure 1). The initial search yielded 93 studies. Based on the titles and abstracts, we excluded 78 studies. We downloaded the remaining studies for full-text review and included the 7 studies that matched the inclusion criteria.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart outlining the study selection process.



### Authors, Year of Publication, and Country of Origin

The included studies were conducted in the United States (5/7) and Korea (2/7) and were published between 2015 and 2021. Birnbaum and colleagues conducted 3 of the 7 included studies.

### Study Design, PRISMA Quality Assessment, and Sample Size

The included studies had a retrospective design. The quality of the studies was assessed using the PRISMA criteria (Table 1 and Multimedia Appendix 4). The quality of the 7 studies was heterogeneous, with a mean PRISMA score of 32.3. Risk of bias varied across the studies. The main bias was a classification bias in 5 of the 7 studies not using a clinical diagnosis. The sample sizes were mostly small, with 6 of the 7 studies having a sample size smaller than 5392 participants. The samples size varied between 51 and 265,396 participants.

**Table 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) quality scores of the included studies.

| Study | PRISMA quality score[a] |
|---|---|
| Birnbaum et al [31], 2017, United States | 31 |
| Kim et al [32], 2020, Korea | 31 |
| Birnbaum et al [33], 2019, United States | 37 |
| Birnbaum et al [34], 2020, United States | 36 |
| McManus et al [35], 2015, United States | 29 |
| Bae et al [36], 2021, Korea | 36 |
| Mitchell et al [37], 2015, United States | 29 |

[a] The higher the score, the better the overall quality.

### Social Media and Technologies Used

Several AI technologies have been used (Table 2). The 2 most commonly used algorithms were SVM and random forest. None of the included studies used data augmentation. Cross-validation techniques to prevent overfitting were used in 5 studies. One study used a dropout rate of 0.25 to prevent overfitting. The social media platforms the studies used were Facebook, Twitter, and Reddit. The studies used mainly linguistic features, as well as some activity-related features (Table 3).

**Table 2.** Information extracted from the included studies.

| Authors, year, and country | Overview and inclusion criteria | Objective | Method | Social media and AI[a] technology | Outcome | Main limitations |
|---|---|---|---|---|---|---|
| Birnbaum et al [31], 2017, United States | Users with a self-disclosed diagnosis of schizophrenia on Twitter between 2012 and 2016. Authors randomly selected 671 users diagnosed with schizophrenia from the primary data set. The control group comprised a random sample of Twitter users collected from individuals without any mentions of "schizophrenia" or "psychosis" in their timeline. | To explore the utility of social media as a viable diagnostic tool in identifying individuals with schizophrenia | Twitter profiles from the training data set were reviewed by 2 physicians to determine the probability of belonging to a patient with schizophrenia. The users were then classified into 3 groups: "yes," "maybe," or "no." The machine learning model was then tested on unseen data of 100 users and its results were compared to those of the 2 physicians. | Twitter. Several algorithms including the Gaussian naïve Bayes (NB), random forest (RF), logistic regression (LR), and support vector machine (SVM) were trained. The best performing algorithm on cross-validation was selected (RF) using 10-fold-cross-validation. | RF yielded an area under the curve (AUC) of 0.88. | The research team only had access to publicly available Twitter profiles. The clinical diagnosis of the included users was unknown. |
| Kim et al [32], 2020, Korea | Data from 228,060 users with 488,472 posts from January 2017 to December 2018 were employed for the analysis. | Aimed to answer the following question: Can we identify whether a user's social media post can be associated with a mental illness? | Collection of post data on mental health–related subreddit groups. The study collected information from 248,537 users, who wrote 633,385 posts in the 7 subreddits. After removal of deleted users and posts, 488,472 posts were analyzed. Authors created 6 models for each mental disorder. Each model was created with the posts of the associated subreddit group. | Reddit. Extreme gradient boosting (XGBoost) and convolutional neural network (CNN) were employed. A dropout rate of 0.25 was used to prevent overfitting issues. | In the schizophrenia subreddit (r/schizophrenia), accuracies of XGBoost and CNN were 86.75% and 94.33%, respectively. | The clinical diagnosis of included subjects was unknown. |
| Birnbaum et al [33], 2019, United States | Participants aged 15 to 35 years diagnosed with a primary psychotic disorder were screened for eligibility. Recruitment occurred between March 2016 and December 2018, and 51 of the included participants had a recent onset of psychosis. | To identify and predict early relapse warning signs in social media activity collected from a cohort of individuals receiving psychiatric care for schizophrenia and other primary psychotic disorders | The authors collected 52,815 Facebook posts across 51 participants with a recent onset of psychosis and applied anomaly detection to explore linguistic and behavioral changes associated with psychotic relapse. | Facebook. Three 1-class SVM models for 3 different data configurations (3 different time periods: 1 month, 2 months, and 3 months). The 1-month period showed the highest specificity, which led to an ensemble 1-class SVM algorithm. | The ensemble model had the highest specificity (0.71) but low sensitivity (0.38). The 3-month model had good sensitivity (0.9) but low specificity (0.04). | Monthly periods of relative health and relative illness were characterized. The illness trajectory of psychotic disorder does not fall only into 2 distinct categories, as the symptoms can fluctuate over time. |

| Authors, year, and country | Overview and inclusion criteria | Objective | Method | Social media and AI[a] technology | Outcome | Main limitations |
|---|---|---|---|---|---|---|
| Birnbaum et al [34], 2020, United States | A total of 3,404,959 Facebook messages and 142,390 images across 223 participants with schizophrenia spectrum disorders (SSD), mood disorders (MD), and healthy volunteers (HV) were collected. Participants aged between 15 and 35 years were recruited from Northwell Health's psychiatry department. | To evaluate whether it was possible to distinguish among SSD, MD, and HV based on Facebook data alone. | The authors analyzed features uploaded up to 18 months before the first hospitalization using machine learning and built classifiers that distinguished SSD and MD from HV as well as SSD from MD. | Facebook. RF and 5-fold cross-validation were used. | Classification achieved AUC values of 0.77 (HV vs MD), 0.76 (HV vs SSD), and 0.72 (SSD vs MD). | Data from Facebook were retrospectively collected. |
| McManus et al [35], 2015, United States | The cohort contained Twitter users who self-identified as having schizophrenia (cases) and users who did not self-identify as having any mental disorder (controls), with 96 cases and 200 controls. A user was defined as a case if 2 or more of the following held true: The user self-identifies in the user description; the user self-identifies in their status updates; the user follows @schizotribe, a known Twitter community of users with schizophrenia. | To distinguish individuals with schizophrenia from control individuals using Twitter data | To distinguish Twitter users with schizophrenia from controls, the authors extracted a set of features from each user's profile and posting history (28 numerical features). | Twitter. Several models: NB, artificial neural networks (ANNs), and SVMs. 5-fold cross validation on the training data. In addition to the raw feature vectors, the authors tested 2 transformations of the feature vectors for each of the models: log scaling of the delay between tweets and principal component analysis (PCA). | The best performing model was an SVM with PCA-transformed features (accuracy of 0.893). The 2 best performing models based on the F1 score involved PCA-transformed features. | Users self-identified as patients with schizophrenia. |
| Bae et al [36], Korea, 2021 | A large corpus of social media posts was collected from web-based Reddit subcommunities for schizophrenia (n= 13,156) and control groups (n=247,569) comprising non-mental health–related subreddits (fitness, jokes, meditation, parenting, relationships, and teaching). | To determine whether machine learning could be effectively used to detect signs of schizophrenia in social media users by analyzing their social media texts | Authors collected posts from subreddit. They only included original posts and excluded the comments. They collected titles and bodies of posts along with user IDs. This resulted in 60,009 original schizophrenia posts from 16,462 users as well as 425,341 posts of the control group from 248,934 users. | Reddit. Posts from the control group were randomly downsampled to create a balanced data set (n= 13,156 posts for each group). The authors evaluated 4 different algorithms, namely SVM, LR, NB, and RF, with 10-fold cross-validation. | AUC values were as follows: RF 0.97, SVM 0.91, LR 0.9, and NB 0.87 | The authors do not have evidence that users of r/schizophrenia are clinically diagnosed. |

| Authors, year, and country | Overview and inclusion criteria | Objective | Method | Social media and AI[a] technology | Outcome | Main limitations |
|---|---|---|---|---|---|---|
| Mitchell et al [37], United States, 2015 | A corpus of users diagnosed with schizophrenia was collected from publicly available Twitter data, including 174 users with an apparently genuine self-stated diagnosis of a schizophrenia-related condition. Random Twitter users were included as the control, and there were equal numbers of users with schizophrenia and community controls. | To examine how linguistic signals may aid in identifying and getting help to people with schizophrenia | Each self-stated diagnosis included in this study was examined by an author to verify that it appeared to be a real statement of a schizophrenia diagnosis, excluding jokes, quotes, or disingenuous statements. For each user, the authors obtained a set of their public Twitter posts via the Twitter application programming interface, collecting up to 3200 tweets. | Twitter. The authors used 10-fold cross-validation and 2 machine learning methods, namely SVM and maximum entropy. | The SVM model reached an 82.3% accuracy. | Clinical diagnosis was unknown. |

[a] AI: artificial intelligence.

**Table 3.** Features used in the included studies.

| Authors, year, and country | Features |
|---|---|
| Birnbaum et al [31], 2017, United States | The authors employed feature scaling to standardize the range of features. The LIWC[a] features were within a normalized range of 0 to 1. The n-gram features represented frequency counts that required standardization. The min-max rescaling technique was used to scale the n-gram features to the range of 0 to 1. They employed feature selection methods to eliminate noisy features. The filter method was used where features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable. Adopting the ANOVA F test reduced the feature space from 550 features to k – best features (where k=350) by removing noisy and redundant features. |
| Kim et al [32], 2020, Korea | The natural language toolkit was implemented in Python (Python Software Foundation) to tokenize users' posts and filter frequently employed words (stop words). Porter stemmer (a tool used to explore word meaning and source) was employed on the tokenized words to convert a word to its root meaning and to decrease the number of word corpora. |
| Birnbaum et al [33], 2019, United States | Facebook timeline data grounded in the symptomatic and functional impairments associated with psychotic disorders were used. These include 3 types of features. The first was word usage and psycholinguistic attributes related to affective, social, and personal experiences. The second included linguistic structural attributes, such as complexity, readability, and repeatability related to thought organization and cognitive abilities. The third comprised web-based activities relating to social functioning and diurnal patterns (friending, posting, and check-ins). |
| Birnbaum et al [34], 2020, United States | Image and linguistic features were used. |
| McManus et al [35], 2015, United States | Features for describing emoticon use and schizophrenia-related words were used. The authors used the natural language toolkit in Python to perform tokenization and lemmatization, before extracting textual features and NumPy for generating the final numeric feature vectors. The final 28 numerical features included the number of Twitter followers, number of followed users, proportion of tweets using schizophrenia-related words, emoticon usage, posting time of day, and posting rate. Two transformations of the feature vectors for each of the models were used: log scaling of the delay between tweets and principal component analysis. |
| Bae et al [36], Korea, 2021 | The linguistic features were extracted using the LIWC package and the *liwcalike* function from the *quanteda* package. Structural and psychological components of the text based on psychometrically validated dictionary, word stems, and emotions assigned to a range of categories were assessed. There were 22 LIWC features for each post: linguistic processes (word count and words more than 6 letters), function words (personal pronouns, first-person singular, first-person plural, second person, third-person singular, third-person plural, and impersonal pronouns), time orientations (past focus, present focus, and future focus), and psychological processes (positive emotion, negative emotion, anger, fear, joy, disgust, sadness, anticipation, trust, and surprise). Linguistic features between the schizophrenia and the control (nonschizophrenia) groups were compared. The D'Agostino and Pearson's test ($\alpha$=.05) were conducted to test whether each of the linguistic features was normally distributed. As data followed a normal distribution, a 2-tailed *t* test was performed to determine whether the linguistic features differed between groups. The threshold of statistical significance was adjusted using the false discovery rate method to correct for multiple comparisons, with *P*<.05 in all cases. |
| Mitchell et al [37], United States, 2015 | All natural language processing features were either automatically constructed or unsupervised, meaning that no manual annotation is required to create them. It is important to note that although these features were inspired by the literature on schizophrenia, they were not direct correlates of standard schizophrenia markers. The authors used the following methods to extract features: perplexity (ppl), Brown-Cluster Dist, LIWC, CLM[b], LIWC+CLM, LDA[c] Topic Dist (TDist), CLM+TDist+BDist+ppl, CLM+TDist, and LIWC+TDist. The authors used the LIWC approach to map the words to psychological concepts as well as open-vocabulary approaches such as LDA, Brown clustering, CLM, or perplexity in order to extract features from the corpus in an unsupervised manner. In particular, the LDA algorithm learns a probability distribution over topics for each document. The Brown clustering is a hierarchical clustering algorithm that groups words that occur in similar contexts. Regarding the CLM method, the idea is to assign a probability to a sequence of words (n-grams). In the paper, the authors used a sequence of 5 characters (5-grams). Finally, perplexity is a measurement of how predictable the language is. We expect a high perplexity score for a user using a noncoherent language. |

[a] LIWC: linguistic inquiry and word count.

[b] CLM: character language model.

[c] LDA: latent Dirichlet allocation.

## Study Objectives and Algorithm Performance

### Main Results

Most studies aimed to identify users with schizophrenia on social media. One study aimed to identify and predict early relapse after hospitalization for schizophrenia [33].

The results were informed by multiple parameters, including the area under the curve (AUC), accuracy, as well as sensitivity and specificity. The AUC of the included studies ranged from 0.76 to 0.97 (Table 4), which is considered to be good to excellent. However, only 1 of these studies [34] used data from clinically diagnosed patients, obtaining an AUC of 0.76. The studies whose results were informed by the accuracy parameter obtained an accuracy ranging from 81% to 96%. One study [33] reported results obtained using predictive models with a sensitivity/specificity couple (Table 5). This study sought to identify and predict relapse of schizophrenia. The authors collected Facebook data from a small sample of patients diagnosed with schizophrenia who had a relapse in the following months. They used these data to build a machine learning model that could be used to analyze the patients' data in real time. They obtained several sensitivity/specificity couples. The 3-month ensemble model showed good sensitivity (90%) although the specificity was low (40%). This is an example of the unique tools that could be developed using AI. A high-sensitivity tool could allow physicians to detect a relapse earlier and offer timely care to their patients. The 1-month model had a high specificity (0.71) but low sensitivity (0.38).

**Table 4.** Performance of the different algorithms in terms of the area under the curve.

| Study | Support vector machine | Random forest | Logistic regression | Naïve Bayes |
|---|---|---|---|---|
| Birnbaum et al [31], 2017, United States | —[a] | 0.88 | — | — |
| Birnbaum et al [34], 2020, United States | — | 0.76 | — | — |
| Bae et al [36], Korea, 2021 | 0.91 | 0.97 | 0.90 | 0.87 |

[a] Not applicable.

**Table 5.** Performance of the different algorithms in terms of accuracy and sensitivity/specificity.

| Study | Accuracy (%) | Sensitivity/specificity (%) |
|---|---|---|
| Birnbaum et al [31], 2017, United States | 81 (RF[a]) | —[b] |
| Kim et al [32], 2020, Korea | 86.75 (XGB[c]), 94.33 (CNN[d]) | — |
| Birnbaum et al [33], 2019, United States | — | 38/71, 90/40 (SVM[e]) |
| McManus et al [35], 2015, United States | 89.3 (SVM) | — |
| Bae et al [36], Korea, 2021 | 86 (NB[f]), 89 (LR[g]), 91 (SVM), 96 (RF) | — |
| Mitchell et al [37], 2015, United States | 82.3 (SVM) | — |

[a] RF: random forest.

[b] Not applicable.

[c] XGB: extreme gradient boosting.

[d] CNN: convolutional neural network.

[e] SVM: support vector machine.

[f] NB: naïve Bayes.

[g] LR: logistic regression.

### *Data Used*

Most of the studies did not have access to clinical diagnostic or health data. Instead, they used evaluations of users' profiles by psychiatrists to access which user could be classified as having schizophrenia. The included studies used the content of the posts to train and test the models. They also used activity-derived markers such as friending, check-ins, and the number of followers. Used features were chosen to represent the symptoms of schizophrenia described in the literature and were focused on identifying disorganized symptoms and cognitive abilities (Table 2).

## Discussion

### Principal Results

On a statistical basis, the included studies reported good to excellent performance. Indeed, many of the metrics they reported on are at the top of their respective ranges (AUC, accuracy, sensitivity, and specificity). However, accuracy has high chances of being biased, as most studies did not have access to clinical diagnostic data to train the models. Most studies used the evaluations of social media profiles by trained physicians to classify patients into different groups. Thus, we cannot reach any conclusions regarding the performance of AI in detecting patients with schizophrenia on social media. Moreover, 2 of the included studies that did have access to clinical diagnostic data [33,34] showed the most conservative results (Tables 3 and 4). The included studies were heterogeneous, and some of them introduced interesting new perspectives. After the first psychotic episode, AI and social media could be used to monitor the clinical state of the patients and detect a relapse sooner. This strategy has been studied by Birnbaum et al [33]. We also hypothesize that it could be useful with cohorts of ultra-high-risk patients. Social media provides a constant flow of data, which could in theory allow for the monitoring of large patient cohorts and detect early signs of a psychotic episode. This tool could be integrated in the care of these patients with their consent.

## Critical Assessment of the Machine Learning Methodology in the Included Studies

The problem we are considering in this review is a binary classification problem (whether a user profile on social media indicates schizophrenia) mainly based on textual data (the user's posts). In this section, we introduce the key challenges that we need to tackle given the collected data, analyze the methodology used in the literature, and present our machine learning methodology to solve this problem.

### Key Challenges in Performing Statistical Studies

The first major obstacle statisticians need to tackle is the imbalanced data set [38]. Indeed, in the included studies, the number of included controls is several times higher than the included cases. Thus, a "naïve" binary classifier (random forest, SVM, logistic regression, etc) should not be used. It would tend to overestimate the dominant class over the minor one. Moreover, the small sample size in the included studies suggests that there is a chance that the probability distribution in clinical practice is different than the one in the training set used for the experiments.

The second challenge imposed by the data structure is the textual data. Indeed, unlike most problems, we must deal with unstructured data, as opposed to structured data where the features are well organized in a table. Here, the data we consider include text (posts of the users), and it is unstructured. Therefore, natural language processing (NLP) techniques are called for to extract relevant features to run a machine learning classifier. This is often a delicate stage, as researchers often introduce an inductive bias when they decide which features to extract from the data. State-of-the art NN algorithms like the BERT (Bidirectional Encoder Representations from Transformers) algorithm [39] can automatically extract features from textual data in a "pretrained stage." The algorithm is then trained to perform the desired classification task during the "fine-tuned stage."

### Analysis of the Methodology Used in Previous Works

Most previous works introduced in the last section extract their features using NLP modules such as the natural language toolkit [32,35], the linguistic inquiry and word count package [36], or even older methods like the n-gram [31]. The major drawback of these "bag-of-words" techniques [40] or term frequency–inverse document frequency methodologies [41] is that they often vectorize the textual data only based on the words (and their statistics in the sentences) without accounting for grammar and semantic relations between them and their context. Thus, these feature extraction methods fall short when capturing semantic or syntactic information or the sentiments of words [42].

### Perspectives for Future Studies: Our Recommendations for Machine Learning Methodology

Based on the aforementioned challenges and the limitations of previous works, we present in this section our approach to identify patients with schizophrenia based on social media activity.

First, we need to address the imbalanced data set problem. Among the numerous approaches, we selected one of the most used methods in practice called SMOTE (Synthetic Minority Oversampling Technique) [43]. The idea is to balance the data set by creating synthetic samples from the minority class so that both classes become more balanced. Specifically, SMOTE selects an example from the minority class and its k neighbors (typically k=5) and creates a synthetic example as the convex combination of these 2 data points. This procedure can produce as many synthetic samples for the minority class as needed, and it guarantees that these created examples are realistic, as they are close to the existing ones in the feature space.

However, for this data augmentation technique to work, the feature space needs to be continuous, which is not the case in textual data. To alleviate this issue, we need to use word embeddings. In NLP, word embedding is a continuous representation of a word that encodes the meaning of the word in a feature space, and it is usually a real-valued vector [44]. Thus, 2 words with close meanings like "ill" and "disease" will be closer in the feature space than 2 words like "ill" and "car." Nevertheless, the remaining challenge is to create a relevant embedding space. A popular method in NLP is to use a pretrained language model like BERT [39]. This algorithm has been pretrained by Google on the concatenation of the two largest data sets: BookCorpus [45], gathering 11,038 unpublished books, and English Wikipedia, gathering 6,427,217 articles. These high-quality embeddings not only allow us to use the SMOTE method to augment the minority class, but they also allow us to represent the textual data in an informative feature space. The latter will be used as the input to our classification algorithm.

Finally, we need to define the binary classifier for our problem. As we have already used the BERT algorithm to create the word embeddings of the posts, the natural approach would be to use it as a binary classifier as well. To do so, we need to "fine-tune" it using our own data set by adding a linear layer to the existing NN architecture. This approach to text classification has demonstrated state-of-the-art results on 8 widely studied text classification data sets [46].

### Limitations of This Study

We performed a systematic review of studies using machine learning to identify schizophrenia on social media. Based on our hypothesis, the main limitation of our review is the small number of included studies. When submitting this paper for publication, the published studies on this subject were limited and we were not aware of any new study that met the inclusion criteria. Our review aimed to evaluate the potential of this technology as a new tool for the care of patients with schizophrenia. Therefore, we used broad inclusion criteria to include more papers. The included studies do not have the same objective and thus, their results cannot be compared. However, they describe the various uses of this technology.

Most of the studies (5/7) did not use clinical diagnostic data. Instead, they used evaluations of the mental states of the included subjects based on their public profile history and the contents of their posts. It is unlikely that this method is efficient in accurately identifying patients with schizophrenia. Future studies should use health data including medical diagnostic data to develop an accurate model.

Furthermore, some of the included studies showed limitations in their methodology and choice of machine learning algorithms. We analyzed these limitations and used them to propose recommendations for future projects.

### Ethical Reflection and Privacy

Machine learning tools could be useful in several ways to improve the care of patients. We could monitor the social media activity of patients to detect psychotic relapses sooner. These tools could also be used to detect the first psychotic episode sooner in patients monitored for high risk of psychosis.

The use of machine learning to predict mental health disease raises ethical questions. In what context should we use such tools? Patients would need to comply with the use of these tools and their data in their care. The use of machine learning would be appropriate only if patients consent to it. Furthermore, many countries are establishing a regulatory framework on AI usage [47,48]. AI tools will have to comply with regulation laws to be used in a clinical setting.

In particular, machine learning algorithms need to be trained on massive amounts of unbiased data. To prevent third parties from using these technologies for other purposes, ensuring the safety of medical data is essential.

### Conclusions

AI brings new perspectives in research on schizophrenia. It could be used to monitor the clinical condition of patients at risk of psychosis or to detect relapses of schizophrenia by observing patients on social media. There are currently only a few studies published on this subject, and most of them do not accurately estimate the potential of this technology. However, this technology could become a new tool in the care for patients with schizophrenia, ultimately reducing the burden on caregivers. It should be developed and used in accordance with ethical and legal frameworks.

---

## Conflicts of Interest

None declared.

---

Multimedia Appendix 1
Machine learning parameters.
[DOCX File , 14 KB - jmir_v24i9e36986_app1.docx ]

---

Multimedia Appendix 2
Search strategy.
[DOCX File , 13 KB - jmir_v24i9e36986_app2.docx ]

---

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 31 KB - jmir_v24i9e36986_app3.docx ]

---

Multimedia Appendix 4
A summary of the included studies and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) certainty assessment.
[XLSX File (Microsoft Excel File), 26 KB - jmir_v24i9e36986_app4.xlsx ]

---

## References

1. 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;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Schizophrenia. World Health Organization. 2022 Jan 10. URL: https://www.who.int/news-room/fact-sheets/detail/schizophrenia [accessed 2022-05-15]
3. Charlson F, Ferrari A, Santomauro D, Diminic S, Stockings E, Scott JG, et al. Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. Schizophr Bull 2018 Oct 17;44(6):1195-1203 [FREE Full text] [doi: 10.1093/schbul/sby058] [Medline: 29762765]
4. Cloutier M, Sanon Aigbogun M, Guerin A, Nitulescu R, Ramanakumar AV, Kamat SA, et al. The economic burden of schizophrenia in the United States in 2013. J Clin Psychiatry 2016 Apr;77(06):764-771. [doi: 10.4088/jcp.15m10278]
5. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019 Mar;19(1):64 [FREE Full text] [doi: 10.1186/s12874-019-0681-4] [Medline: 30890124]
6. Deo RC. Machine learning in medicine. Circulation 2015 Nov;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]

7. Lin E, Lin C, Lane H. Precision psychiatry applications with pharmacogenomics: artificial intelligence and machine learning approaches. Int J Mol Sci 2020 Feb;21(3):969 [FREE Full text] [doi: 10.3390/ijms21030969] [Medline: 32024055]
8. Mothi SS, Sudarshan M, Tandon N, Tamminga C, Pearlson G, Sweeney J, et al. Machine learning improved classification of psychoses using clinical and biological stratification: Update from the bipolar-schizophrenia network for intermediate phenotypes (B-SNIP). Schizophr Res 2019 Dec;214:60-69. [doi: 10.1016/j.schres.2018.04.037] [Medline: 29807804]
9. Lejeune A, Le Glaz A, Perron P, Sebti J, Baca-Garcia E, Walter M, et al. Artificial intelligence and suicide prevention: a systematic review. Eur Psychiatry 2022 Feb;65(1):1-22 [FREE Full text] [doi: 10.1192/j.eurpsy.2022.8] [Medline: 35166203]
10. Ma Y, Ji J, Huang Y, Gao H, Li Z, Dong W, et al. Implementing machine learning in bipolar diagnosis in China. Transl Psychiatry 2019 Nov;9(1):305 [FREE Full text] [doi: 10.1038/s41398-019-0638-8] [Medline: 31740657]
11. Gao S, Calhoun VD, Sui J. Machine learning in major depression: from classification to treatment outcome prediction. CNS Neurosci Ther 2018 Nov;24(11):1037-1052 [FREE Full text] [doi: 10.1111/cns.13048] [Medline: 30136381]
12. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annu Rev Public Health 2018 Apr;39:95-112. [doi: 10.1146/annurev-publhealth-040617-014208] [Medline: 29261408]
13. Mori K, Haruno M. Differential ability of network and natural language information on social media to predict interpersonal and mental health traits. J Pers 2021 Apr;89(2):228-243 [FREE Full text] [doi: 10.1111/jopy.12578] [Medline: 32654146]
14. Rezaii N, Walker E, Wolff P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. NPJ Schizophr 2019 Jun;5(1):9 [FREE Full text] [doi: 10.1038/s41537-019-0077-9] [Medline: 31197184]
15. Le Glaz A, Haralambous Y, Kim-Dufor D, 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;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
16. Tariq S, Akhtar N, Afzal H, Khalid S, Mufti MR, Hussain S, et al. A novel co-training-based approach for the classification of mental illnesses using social media posts. IEEE Access 2019 Nov;7:166165-166172. [doi: 10.1109/ACCESS.2019.2953087]
17. Ophir Y, Tikochinski R, Asterhan C, Sisso I, Reichart R. Deep neural networks detect suicide risk from textual facebook posts. Sci Rep 2020 Oct 07;10(1):16685 [FREE Full text] [doi: 10.1038/s41598-020-73917-0] [Medline: 33028921]
18. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, et al. Characterisation of mental health conditions in social media using informed deep learning. Sci Rep 2017 Mar;7(1):45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
19. Kriegeskorte N, Golan T. Neural network models and deep learning. Curr Biol 2019 Apr;29(7):R231-R236 [FREE Full text] [doi: 10.1016/j.cub.2019.02.034] [Medline: 30939301]
20. Hearst M, Dumais S, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Appl 1998 Jul;13(4):18-28. [doi: 10.1109/5254.708428]
21. Bey R, Goussault R, Grolleau F, Benchoufi M, Porcher R. Fold-stratified cross-validation for unbiased and privacy-preserving federated learning. J Am Med Inform Assoc 2020 Aug 01;27(8):1244-1251 [FREE Full text] [doi: 10.1093/jamia/ocaa096] [Medline: 32620945]
22. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. JMLR 2014;14:1929-1958 [FREE Full text] [doi: 10.1017/cbo9781107298019.021]
23. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010 Sep;5(9):1315-1316 [FREE Full text] [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]
24. Sasaki Y. The truth of the F-measure. 2007 Oct 26. URL: https://www.toyota-ti.ac.jp/Lab/Denshi/COIN/people/yutaka.sasaki/F-measure-YS-26Oct07.pdf [accessed 2022-07-26]
25. Kaggle. URL: https://www.kaggle.com/lct14558/imbalanced-data-why-you-should-not-use-roc-curve [accessed 2022-07-26]
26. Látalová K. Violence and duration of untreated psychosis in first-episode patients. Int J Clin Pract 2014 Mar;68(3):330-335. [doi: 10.1111/ijcp.12327] [Medline: 24471741]
27. Albert N, Weibell MA. The outcome of early intervention in first episode psychosis. Int Rev Psychiatry 2019 Aug;31(5-6):413-424. [doi: 10.1080/09540261.2019.1643703] [Medline: 31456455]
28. Jo YT, Joo SW, Shon S, Kim H, Kim Y, Lee J. Diagnosing schizophrenia with network analysis and a machine learning method. Int J Methods Psychiatr Res 2020 Mar;29(1):e1818 [FREE Full text] [doi: 10.1002/mpr.1818] [Medline: 32022360]
29. de Boer JN, Brederoo SG, Voppel AE, Sommer IE. Anomalies in language as a biomarker for schizophrenia. Curr Opin Psychiatry 2020 May;33(3):212-218. [doi: 10.1097/YCO.0000000000000595] [Medline: 32049766]
30. Feldman J, Hamlyn A, Rice T. Social media in screening and monitoring for early intervention in psychosis. Schizophr Res 2021 Dec;238:70-72. [doi: 10.1016/j.schres.2021.09.023] [Medline: 34607256]
31. Birnbaum ML, Ernala SK, Rizvi AF, De CM, Kane JM. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. J Med Internet Res 2017 Aug;19(8):e289 [FREE Full text] [doi: 10.2196/jmir.7956] [Medline: 28807891]
32. Kim J, Lee J, Park E, Han J. A deep learning model for detecting mental illness from user content on social media. Sci Rep 2020 Jul;10(1):11846 [FREE Full text] [doi: 10.1038/s41598-020-68764-y] [Medline: 32678250]

33. Birnbaum ML, Ernala SK, Rizvi AF, Arenare E, R Van Meter A, De Choudhury M, et al. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. NPJ Schizophr 2019 Oct;5(1):17 [FREE Full text] [doi: 10.1038/s41537-019-0085-9] [Medline: 31591400]
34. Birnbaum ML, Norel R, Van Meter A, Ali AF, Arenare E, Eyigoz E, et al. Identifying signals associated with psychiatric illness utilizing language and images posted to Facebook. NPJ Schizophr 2020 Dec;6(1):38 [FREE Full text] [doi: 10.1038/s41537-020-00125-0] [Medline: 33273468]
35. McManus K, Mallory EK, Goldfeder RL, Haynes WA, Tatum JD. Mining Twitter data to improve detection of schizophrenia. AMIA Jt Summits Transl Sci Proc 2015 Mar;2015:122-126 [FREE Full text] [Medline: 26306253]
36. Bae YJ, Shim M, Lee WH. Schizophrenia detection using machine learning approach from social media content. Sensors (Basel) 2021 Sep;21(17):5924 [FREE Full text] [doi: 10.3390/s21175924] [Medline: 34502815]
37. Mitchell M, Hollingshead K, Coppersmith G. Quantifying the Language of Schizophrenia in Social Media. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology; June 5, 2015; Denver, CO p. 11-20. [doi: 10.3115/v1/w15-1202]
38. Haibo He, Garcia E. Learning from imbalanced data. IEEE Trans Knowl Data Eng 2009 Sep;21(9):1263-1284. [doi: 10.1109/TKDE.2008.239]
39. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv Preprint posted online Oct 11, 2018 [FREE Full text]
40. Zhang Y, Jin R, Zhou ZH. Understanding bag-of-words model: a statistical framework. Int J Mach Learn & Cyber 2010 Aug 28;1(1-4):43-52 [FREE Full text] [doi: 10.1007/s13042-010-0001-0]
41. Leskovec J, Rajaraman A, Ullman JD. Mining of Massive Datasets, 2nd Edition. Cambridge, UK: Cambridge University Press; 2014.
42. Naseem U, Razzak I, Khan S, Prasad M. A Comprehensive Survey on Word Representation Models: From Classical to State-Of-The-Art Word Representation Language Models. arXiv Preprint posted online Oct 8, 2020 [FREE Full text] [doi: 10.1145/3434237]
43. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. JAIR 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
44. Jurafsky D, Martin J. Speech and Language Processing, 2nd Edition. United States: Prentice Hall; 2008.
45. Zhu Y, Kiros R, Zemel R. Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books. arXiv Preprint posted online Jun 22, 2015 [FREE Full text] [doi: 10.1109/iccv.2015.11]
46. Sun C, Qiu X, Xu Y, Huang X. How to Fine-Tune BERT for Text Classification? Chinese Computational Linguistics. Vol 11856. Lecture Notes in Computer Science 2019:194-206. [doi: 10.1007/978-3-030-32381-3_16]
47. Horizon Scanning Assessment Report? Artificial Intelligence. ICMRA. 2021 Apr 21. URL: https://www.icmra.info/drupal/sites/default/files/2021-08/horizon_scanning_report_artificial_intelligence.pdf [accessed 2022-05-15]
48. Fostering a European Approach to Artificial Intelligence. European Commission. 2021 Apr 21. URL: https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence [accessed 2022-05-15]

## Abbreviations

**AI:** artificial intelligence
**AUC:** area under the curve
**CT:** computed tomography
**LASSO:** least absolute shrinkage and selection operator
**NLP:** natural language processing
**NN:** neural network
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**SMOTE:** Synthetic Minority Oversampling Technique
**SVM:** support vector machine

---

*Edited by A Mavragani; submitted 02.02.22; peer-reviewed by A Hudon, A Benis; comments to author 20.04.22; revised version received 30.05.22; accepted 07.06.22; published 06.09.22.*

*Please cite as:*
*Lejeune A, Robaglia BM, Walter M, Berrouiguet S, Lemey C*
*Use of Social Media Data to Diagnose and Monitor Psychotic Disorders: Systematic Review*
*J Med Internet Res 2022;24(9):e36986*
*URL: https://www.jmir.org/2022/9/e36986*
*doi:10.2196/36986*
*PMID:36066938*

Review

# Effects of Serious Games on Depression in Older Adults: Systematic Review and Meta-analysis of Randomized Controlled Trials

Yesol Kim[1], MSN; Soomin Hong[1], PhD; Mona Choi[2,3], PhD

[1]College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea
[2]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea
[3]Yonsei Evidence Based Nursing Centre of Korea, A JBI Affiliated Group, Seoul, Republic of Korea

**Corresponding Author:**
Mona Choi, PhD
Mo-Im Kim Nursing Research Institute
College of Nursing
Yonsei University
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3341
Fax: 82 2 2227 8303
Email: monachoi@yuhs.ac

## Abstract

**Background:** Depression is a severe psychological concern that negatively affects health in older adults. Serious games applied in various fields are considered appropriate interventions, especially in mental health care. However, there is a lack of evidence regarding the effects of serious games on depression in older adults.

**Objective:** This study aimed to investigate the characteristics and effectiveness of serious games for depression in older adults.

**Methods:** A systematic review and meta-analysis of randomized controlled trials were conducted. In total, 5 electronic databases (PubMed, CINAHL, Embase, PsycINFO, and Cochrane Library) were searched to identify relevant studies published until July 6, 2021. A total of 2 reviewers independently conducted study selection, data extraction, and quality appraisals. The risk of bias in the included studies was assessed using the JBI Critical Appraisal Checklist. For the meta-analysis, the effect size was calculated as the standardized mean difference (SMD) by using a random effects model.

**Results:** A total of 17 studies with 1280 older adults were included in the systematic review, and 15 studies were included in the meta-analysis. Serious game interventions were classified into 3 types: physical activity (PA), cognitive function, and both PA and cognitive function. The meta-analysis demonstrated that serious games reduced depression in older adults (SMD −0.54, 95% CI −0.79 to −0.29; *P*<.001). Serious games had a more significant effect size in community or home settings (SMD −0.61, 95% CI −0.95 to −0.26; *P*<.001) than in hospital settings (SMD −0.46, 95% CI −0.85 to −0.08; *P*=.02); however, the difference between groups was not significant. Among the types of games, games for PA (SMD −0.60, 95% CI −0.95 to −0.25; *P*<.001) and games for both (SMD −0.73, 95% CI −1.29 to −0.17; *P*=.01) had a significant effect on reducing depression in older adults. However, no significant correlations were observed between the duration or number of serious games and depression.

**Conclusions:** Serious games were beneficial in reducing depression in older adults. Regardless of the study setting, serious games appeared to reduce depression. Particularly, serious games including PA had a significant impact on reducing depression. Furthermore, high-quality randomized controlled trials are needed to establish substantial evidence for the effectiveness of serious games on depression in older adults.

**Trial Registration:** PROSPERO CRD42021242573; https://tinyurl.com/26xf7ym5

***(J Med Internet Res 2022;24(9):e37753)*** doi:10.2196/37753

**KEYWORDS**

effectiveness; serious game; exergaming; video games; virtual reality; depression; older adults; systematic review; meta-analysis; mobile phone

## Introduction

### Background

The aging population is increasing worldwide. The average life span has increased, and health-related concerns in older adults require attention [1], highlighting their physical and mental health concerns [2]. Researchers have examined the psychosocial aspects of older adults, including anxiety, depression, sleep disorders, loneliness, and social functional impairment [3-5]. Specifically, depression is a severe and typical mental health concern characterized by sadness and hopelessness [6]. Older adults are more vulnerable to depression owing to their psychosocial concerns. During the COVID-19 pandemic, older adults had a high risk of depression owing to a decrease in social relations, regardless of their environment or context [7,8].

The clinical conditions of older adults also make them likely to develop depression. Physical and cognitive problems and functional loss are the primary causes of depression in older adults [2]. Older adults experience ambiguous symptom profiles of depression, and atypical symptoms make early detection challenging [9]. In addition, insufficient psychosocial relationships and decreased economic status after retirement lead to depression in older adults [10,11].

The prevention, early detection, and treatment of depression in older adults are crucial. However, older adults tend to avoid using mental health services because of poor physical function, psychological barriers, and reduced mobility [12]. To improve the mental health of older adults, specific methods are required that consider their characteristics and attributes. Moreover, detailed and personalized interventions are required to manage depression in older adults. For instance, physical function, cognitive function (CF), sensory function impairment, comorbidities, medication, and environmental factors should be considered [13-16].

Digital interventions for mental health care are considered promising [17,18] and have become indispensable since the COVID-19 pandemic. Digital interventions have been found to be effective in reducing the symptoms of depression [19], loneliness [5], and social isolation [20] in older adults. There has been a gradual increase in the use of digital interventions for older adults in clinics and research, although digital interventions may pose certain challenges to older adults [21,22]. A mixed methods study has identified that although older adults may wish to make use of digital interventions to alleviate depression, they might also initially face certain obstacles to participation [23]. Digital interventions that consider the daily lives of older adults, ease of use, and low cost may help reduce depression [23]. In light of their perspectives, circumstances, and contexts, it is necessary to develop and implement effective interventions for older adults.

Serious games, a type of digital intervention, refer to a series of activities performed by combining the aspects of video games for specific purposes, such as education or rehabilitation [24]. Initially, they were developed for military purposes in the 1970s and recently appeared in more advanced forms with the development of computers and mobile devices [25]. In addition, they are now widely used in education and health care [26], such as physical rehabilitation [27], cognitive training [28,29], and health promotion [30]. Interest in serious games which allow participants to voluntarily achieve their goals has increased.

Various interventions of serious games have been conducted not only for adolescents [31] and younger adults [32] but also for older adults [33-35]. They were applied to older adults in different ways, including video games, using devices such as Nintendo [36,37], and virtual reality serious games [38]. When developing serious game interventions to improve the health of older adults, there are a variety of goals, such as strengthening physical function or CF. In addition, the composition or content of serious games was altered to fit the purpose of the intervention. For instance, if a serious game is designed to improve physical function, older adults require to move their bodies during the game [39]. In a previous study, a serious game was used and evaluated to enhance spatial memory in older adults by implementing the appearance of a real-world city and systematically applying a virtual environment [40].

Serious games are considered appropriate interventions in mental health care, including for the general population and patients with or without psychiatric concerns [41-43]. Studies have reported that serious games influence depression [44,45]; however, a systematic review and meta-analysis focusing on the effect size of a serious game on depression in older adults is rare. In addition, the intervention effects differed in studies on serious games for older adults [33,46-48]. Therefore, further analysis of serious games for depression in older adults is required.

### Aims

This systematic review and meta-analysis aimed to analyze the effects of various types of serious games on depression in older adults. The detailed research questions leading to this study are as follows: (1) What are the characteristics of serious games used to intervene in depression among older adults? (2) How effective are serious games to intervene in depression among older adults? (3) Which aspects of serious games affect depression in older adults?

## Methods

### Design

This systematic review and meta-analysis of randomized controlled trials were reported following the guidelines of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [49] and the JBI manual [50]. The study protocol was registered with PROSPERO (registration number CRD42021242573).

### Search Strategy

We conducted a systematic search by using electronic databases (PubMed, CINAHL, Embase, PsycINFO, and the Cochrane Library) on July 6, 2021. To organize the search terms, free text and Medical Subject Headings terms were combined according to the participants, interventions, comparisons, and outcomes. Key search terms were as follows: (“aged” or “older” or “elder*” or “senior”) and (“game” or “gaming” or “exergame” or “serious game” or “serious gaming”) and (“depression” or “depressive disorder”). A summary of search strategies is presented in Multimedia Appendix 1.

### Eligibility Criteria

Eligibility criteria were determined according to the participants, interventions, comparisons, outcomes, and study design. The inclusion criteria were as follows: (1) participants—studies that included older adults with a mean age of ≥65 years; (2) interventions—studies that applied serious games comprising exergames, virtual reality games, or digital games; (3) comparisons—studies that applied usual care or nonserious games for the control group; (4) outcomes—studies measuring depression; and (5) study design—randomized controlled trials only. The exclusion criteria were as follows: (1) studies that applied different doses or intensities of the serious game for the control group; (2) studies published in a language other than English; and (3) gray literature such as theses, dissertations, or conference abstracts.

### Study Selection

After searching for studies in electronic databases, one researcher (YK) exported all the studies to the reference management software EndNote X9 (Clarivate Analytics), and the other researchers (SH and MC) rechecked the extracted studies. The titles and abstracts of the extracted studies were independently screened according to the inclusion criteria by 2 researchers (YK and SH). Subsequently, they reviewed the full texts separately to select the final studies to be included. If the screening results did not match, a consensus was reached through discussion. The other researcher (MC) supervised the screening process.

### Data Extraction

Two researchers (YK and SH) independently extracted the data. The data extraction of the selected studies was performed using a structured form that included study, participant, intervention, and outcome characteristics. First, the study characteristics included authors, publication year, country, and setting. Second, the participant characteristics included health status, age (mean and SD), and sample size. Intervention characteristics consisted of type, the device used, content, duration, frequency, time, dose of serious games, type of control group, and the interventionist. Finally, outcome characteristics comprised the measurement of depression, the main result, mean, and SD for the experimental and control groups. If data required for analysis could not be found in the article, researchers requested data from the respective authors via email.

### Risk-of-Bias Assessment

Two researchers (YK and SH) independently evaluated the methodological quality of the included studies according to the JBI Critical Appraisal Checklist for randomized controlled trials [50]. The checklist has 13 items to assess the risk of bias, including participants, assignments, measurement, and analysis domains, and 1 overall appraisal item. After each included study was assessed using a 13-item checklist as “yes,” “no,” and “unclear,” the final quality judgment was drawn according to the “yes” ratio. The risk of bias was evaluated as follows: (1) ≥75% was ranked as high quality, (2) the range between 50% and 74% was ranked as medium quality, and (3) <50% was considered poor quality [51].

### Data Analysis

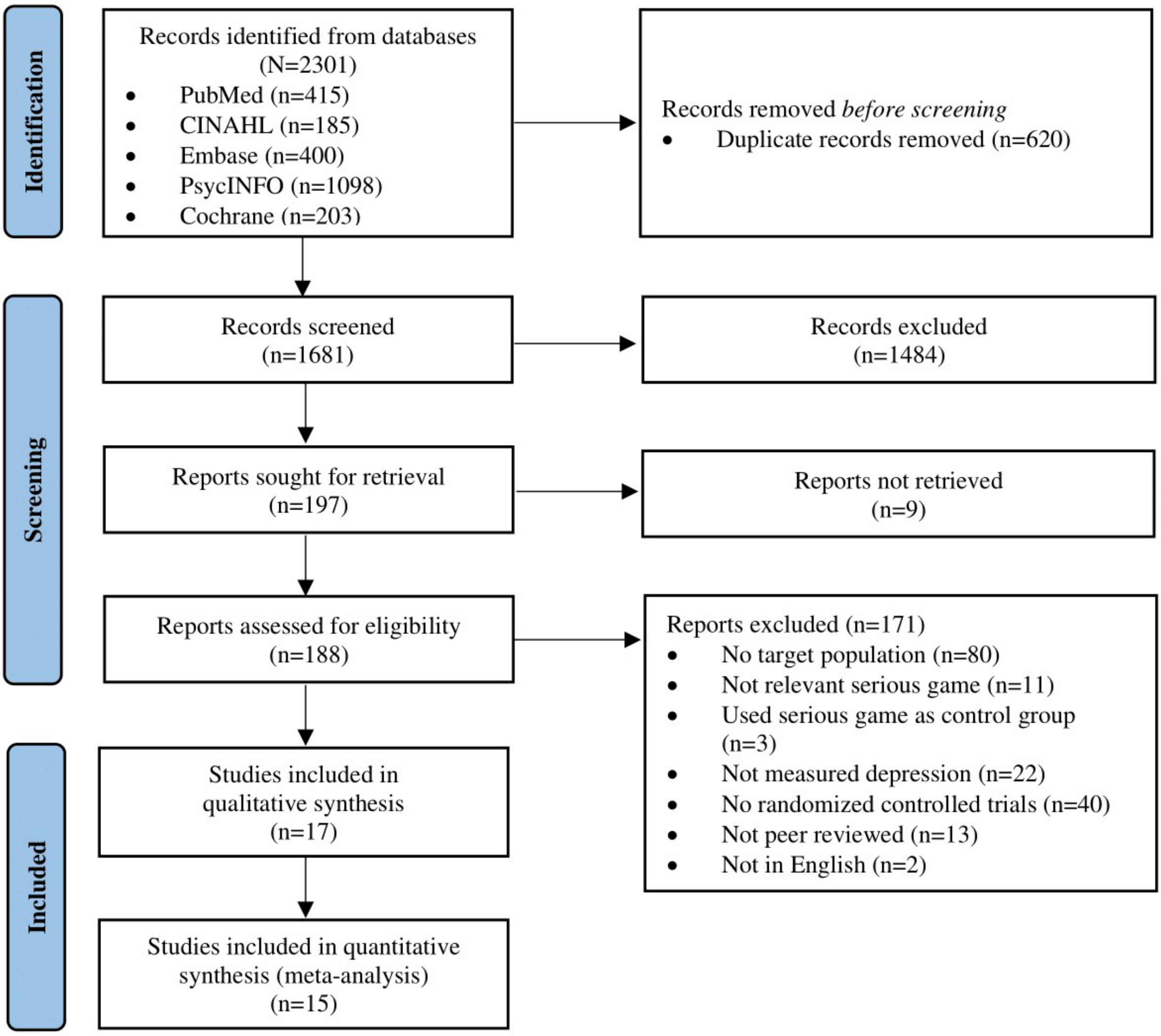
We calculated effect sizes as the standardized mean difference (SMD) with a 95% CI by using the mean and SD to synthesize the pooled effect of serious games on depression in older adults. When depression values were not presented as mean and SD in original studies, SMD was calculated through the conversion process by using SE, median, range, or IQR [52,53]. On the basis of the study by Cohen [54], the effect sizes were considered small (0.2≤SMD<0.5), medium (0.5≤SMD<0.8), and large (SMD≥0.8). To identify the effects of heterogeneity in the meta-analysis, $I^2$ was used. Heterogeneity was classified as low (25%), moderate (50%), or high (75%) according to $I^2$ values [55]. Because the characteristics of the participants and interventions included in the meta-analysis were heterogeneous, we conducted a meta-analysis by using a random effects model. According to the heterogeneity results, a subgroup analysis was conducted on the setting, participant characteristics, type of serious game, and the control group. In addition, we performed a meta-regression to explore the causes of heterogeneity regarding the duration and dose of serious games. Publication bias was assessed using funnel plots and Egger test. Statistical significance was set at *P*<.05. Statistical analysis was performed using the Comprehensive Meta-analysis software (version 3.0) and Review Manager software (version 5.4).

## Results

### Study Selection

Figure 1 shows a PRISMA 2020 flow diagram for screening and selection. A total of 2301 studies were extracted from 5 databases. After removing duplicate records (n=620, 26.94%), 1681 (73.06%) studies were screened using the inclusion and exclusion criteria based on title, abstract, and full text. Finally, 0.74% (17/2301) of studies were included in this study.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for study screening and selection.



## Risk of Bias of the Studies

The results for the risk of bias in this study are presented in Multimedia Appendix 2 [50-66]. As confirmed by the number that met the criteria “yes,” the distribution of satisfying the risk-of-bias evaluation items ranged from 4 (31%) to 10 (77%) out of 13 items. Only 1 study [56] was evaluated as high quality, which indicated a low risk of bias. In total, 5 studies were ranked as poor quality (<50%) [57-61]. The remaining 11 studies were categorized as medium quality, with a distribution of 54% to 69%.

## Study Characteristics

The characteristics of the selected studies are summarized in Table 1. A total of 17 studies were published between 2012 and 2021 and conducted in 10 countries. The United States had the most research conducted [57,58,60,62], followed by Brazil [61,63,64], Korea [59,65], and Hungary [66,67]. More than half of the studies were conducted in community settings such as retirement villages, nursing homes, long-term care facilities, and assisted living facilities [56,60-63,68-71]. In total, 5 studies were conducted in hospitals [59,64-67] and 3 in homes [57,58,72].

A total of 1280 older adults were included in this review. Among the 17 studies, 7 (41%) mentioned health problems as the criteria for participant selection. The health status characteristics of the participants included neurocognitive problems, such as Parkinson disease and predementia [64-66,71], depression [57], limited mobility [67], and stroke [59]. For the 3-arm study, control groups were indicated by “control group a” and “control group b” [58,64,66,67]. The mean age of the participants ranged from 66.4 (SD 0.8) [63] to 85.0 (SD 6.1) [71], and the sample size varied between 16 [69] and 351 [62].

Various measurements of depression in older adults were described in the included studies. In 35% (6/17) of the studies, the Geriatric Depression Scale, including 5, 15, and 30 items, was used to measure depression [58,60,64,65,68,70]. In 29% (5/17) of the studies, depression was measured using the Beck Depression Inventory [59,61,66,67,69]. Other studies (6/17, 35%) used the Patient Health Questionnaire-9 [56,62], Profile of Mood States [63,72], Cornell Scale for Depression in Dementia [71], and Hamilton Depression Rating Scale [57].

**Table 1.** Characteristics of the included studies (N=17).

| Authors, year | Country, setting | Characteristics of participants | Age (years), mean (SD) | Sample size, n | Measure | Effect size[a] (95% CI) |
|---|---|---|---|---|---|---|
| Rendon et al [60], 2012 | United States, community | Community-dwelling older adults | E[b]: 85.7 (4.3); C[c]: 83.3 (6.2) | E: 20; C: 20 | GDS[d]-30 | −0.72 (−1.41 to −0.03) |
| Schoene et al [56], 2015 | Australia, community | Community-dwelling older adults | 81.5 (7.0) | E: 47; C: 43 | PHQ-9[e] | −0.29 (−0.72 to 0.14) |
| Choi et al [59], 2016 | Korea, hospital | Hospitalized patients with ischemic stroke | E: 61.0 (15.2); C: 72.1 (9.9) | E: 12; C: 12 | BDI[f] | −0.11 (−0.91 to 0.69) |
| Levy et al [69], 2016 | France, community | Community-dwelling older adults with the fear of falling | E: 72.4 (12.3); C: 68.7 (19.1) | E: 9; C: 7 | BDI | −0.75 (−1.77 to 0.27) |
| Nouchi et al [72], 2016 | Japan, home | Community-dwelling older adults | 68.9 (3.7) | E: 36; C: 36 | POMS[g]2 | NC[h] |
| Anguera et al [57], 2017 | United States, home | Major depression | 68.0 (6.3) | E: 12; C: 10 | HAM-D[i] | −0.13 (−0.97 to 0.71) |
| Ferraz et al [64], 2018 | Brazil, hospital | Parkinson disease | 69.0 (5.0) | E: 22; C-a[j]: 25; C-b[k]: 25 | GDS-15 | a: −0.10 (−0.71 to 0.51); b: −0.11 (−0.74 to 0.52) |
| Belchior et al [58], 2019 | United States, home | Community-dwelling older adults | 73.2 (5.5) | E: 26; C-a: 20; C-b: 25 | GDS-30 | NC |
| Smith et al [62], 2019 | United States, community | Lived in supported senior living settings | 80.6 (9.1) | E: 173; C: 178 | PHQ-9 | −0.10 (−0.34 to 0.14) |
| Stanmore et al [70], 2019 | United Kingdom, community | Lived in assisted living facilities | E: 77.9 (8.9); C: 77.8 (10.2) | E: 56; C: 50 | GDS-5 | −0.17 (−0.58 to 0.24) |
| Tollár et al [66], 2019 | Hungary, hospital | Parkinson disease | E: 70.0 (4.7); C-a: 70.6 (4.1); C-b: 67.5 (4.3) | E: 25; C-a: 25; C-b: 24 | BDI | a: −0.17 (−0.72 to 0.38); b: −1.22 (−1.83 to −0.61) |
| Tollár et al [67], 2019 | Hungary, hospital | Mobility-limited older adults | 69.6 (3.5) | E: 28; C-a: 27; C-b: 28 | BDI | a: −0.28 (−0.81 to 0.25); b: −1.46 (−2.05 to −0.87) |
| de Morais et al [63], 2020 | Brazil, community | Older adults | 66.4 (0.8) | E: 29; C: 29 | POMS | −0.29 (−0.80 to 0.22) |
| Rica et al [61], 2020 | Brazil, community | Institutionalized older women aged >60 years | Not reported | E: 16; C: 34 | BDI | −2.08 (−2.81 to −1.35) |
| Jahouh et al [68], 2021 | Spain, community | Institutionalized in nursing home or attending day center | E: 85.1 (8.6); C: 83.3 (8.8) | E: 40; C: 40 | GDS-15 | −0.65 (−1.10 to −0.20) |
| Kang et al [65], 2021 | Korea, hospital | Predementia state | 74.5 (5.8) | E: 25; C: 20 | GDS-30 | −0.19 (−0.82 to 0.44) |
| Swinnen et al [71], 2021 | Belgium, community | Older adults with neurocognitive disorder residing in long-term care facilities | E: 84.7 (5.6); C: 85.3 (6.5) | E: 28; C: 27 | CSDD[l] | −1.38 (−2.03 to −0.73) |

[a] Effect size was calculated as the standardized mean difference with a 95% CI.

[b] E: experimental group.

[c] C: control group.

[d] GDS: Geriatric Depression Scale.

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

[f] BDI: Beck Depression Inventory.

[g] POMS: Profile of Mood State.

[h] NC: not calculated because required data were not provided.

[i] HAM-D: Hamilton Depression Rating Scale.

[j] C-a: a control group of the 3-arm study.

[k] C-b: the other control group of the 3-arm study.

[l] CSDD: Cornell Scale for Depression in Dementia.

### Characteristics of the Serious Game Intervention

The characteristics of the serious game intervention are presented in Tables 2 and 3. Serious game interventions were classified into 3 types: games for physical activity (PA; 9/17, 53%), games for CF (5/17, 29%), and games for both PA and CF (3/17, 18%).

Regarding games for PA, the devices used in the intervention were Microsoft Xbox 360 [61,63,64,66,67]; Nintendo [60]; Microsoft Kinect [70]; tablet computers and smartphones with a Bluetooth connection [59]; and V8 Head Mount Display, 3D electromagnetic sensor, and PlayStation 2 [69]. Studies using the Xbox 360 primarily provided various commercial games that involved the movement of the participant's body. Studies using Nintendo had applied strength training, aerobics, and balance games of Wii Fit. A study applying Kinect provided 16 exergames targeting the lower or upper limbs by using the Medical Interactive Recovery Assistant digital platform [70]. Choi et al [59] reported that a game improved the mobility of the upper extremity through a mobile app and Bluetooth connection to smart devices. A study using the V8 Head Mount Display, part of a virtual reality game, provided video games that required movements of the participants' bodies [69].

Regarding games for CF, devices included tablet computers [57,72], controllers [58], CDs or computers [62], and Oculus Lift CV1 and touch controllers [65]. Studies using tablet computers applied the developed cognitive training game to participants. Commercial games that can improve CF were provided based on controllers, CDs, or websites. As part of a virtual reality game, a study using the Oculus Lift CV1 and touch controllers applied games that consisted of multidomain cognitive tasks.

Regarding games for both PA and CF, the devices used were electronic step pads [56,71] and Nintendo [68]. Studies applying electronic step pads provided step training that promoted both PA and CF, wherein participants moved in various directions. Another study provided various games that involved PA and CF using Nintendo (Table 2).

Among the 17 studies, 1 (6%) study [63] did not report the duration of intervention, whereas 16 (94%) studies reported the duration of intervention to be between 2 weeks [59] and 12 months [62]. Among the studies, the most frequent durations of intervention were reported as 8 weeks [57,64,68,71] and 12 weeks [58,61,69,70]. The prescribed serious game intervention was conducted for 1 to 5 sessions per week and 15 to 60 minutes per session. The total dose provided to the participants ranged from 2 sessions [63] to 60 sessions [58].

Of the included studies, the control groups consisted of usual care [58,60,65-70], exercise [59,64,66,67], nonserious games [61,62,72], watching a film or music videos [63,71], or other programs such as providing a brochure [56], cognitive training [58], and problem-solving therapy [57].

Of the 17 studies, 9 (53%) reported interventionists. Physical or occupational therapists [59,60,64,66,67,70,71], a neuropsychologist [65], and an interprofessional team [57] provided interventions to the participants (Table 3).

**Table 2.** Summary of serious game interventions of the included studies (N=17).

| Authors, year | Type of serious game; device | Contents |
|---|---|---|
| Rendon et al [60], 2012 | PA[a]; Nintendo | • Wii fit using the Wii Balance Board<br>• Balance games (lunges, single leg extensions, and twists) |
| Schoene et al [56], 2015 | Both; electronic step pad | • The interactive training system used stepping onto an electronic step pad to interact with a computer interface, and videogame technology was used to deliver the training tasks on standard home television screens<br>• Videogames (Stepper, StepMania, Trail-Stepping, and Tetris) |
| Choi et al [59], 2016 | PA; tablet computer and smartphone with Bluetooth connection | • The MoU-Rehab consisted of 4 mobile game apps<br>• All game apps were designed to improve strength, endurance, range of motion, control, speed, and accuracy of movement in the upper extremity |
| Levy et al [69], 2016 | PA; V8 Head Mount Display, 3D electromagnetic sensor, and EyeToy interface for PlayStation 2 | • Participants played video games that required moving their bodies<br>• Games (wash a window and kung fu) |
| Nouchi et al [72], 2016 | CF[b]; tablet computer | • In total, 12 processing speed training games to function on the tablet computer<br>• All games required participants to detect, identify, discriminate, and localize targets as quickly as possible |
| Anguera et al [57], 2017 | CF; tablet computer | • Mobile iPad intervention called Project: EVO based on the video game called NeuroRacer<br>• This game involves guiding a character through an immersive environment while responding to select targets, with the design format being ideally entertaining |
| Ferraz et al [64], 2018 | PA; Xbox 360 | • Exergames use full-body motion to allow the player to engage in a variety of mini games, all of which feature jump-in, jump-out multiplayer play<br>• Physical components involved in those games included strength and muscular endurance, cardiorespiratory fitness, postural balance, and executive function |
| Belchior et al [58], 2019 | CF; videogame and controller | • Crazy taxi is a driving game with key features that include rapid navigation through an urban environment, attending to speed, and roadway features<br>• Characteristics of this game were speed; elevated perceptual, cognitive, and motor loads; and having items of interest often presented at the periphery of the visual field and under divided attention conditions |
| Smith et al [62], 2019 | CF; CDs or web using computer | • Road Tour on CDs and Double Decision, a web-based version, were used<br>• Road Tour and Double Decision performed the same way<br>• Speed of processing training participants saw an object (either a car or truck) in the center of the monitor and a target (route 66 road sign) along with 7 rabbit distractor signs in a near-periphery orbit. Participants viewed the monitor image as quickly as they could while still correctly identifying the object and the target location |
| Stanmore et al [70], 2019 | PA; Microsoft Kinect | • Kinect tracks the user's performance and records parameters<br>• Each participant was given a prescribed program of standardized exergames that suited the participant's starting level of ability with tailored progression<br>• Individual exercise programs can be tailored using a choice of games for lower or upper limb exercises using 16 of Medical Interactive Recovery Assistant's exergames (strength, balance, coordination, and flexibility exercises) |
| Tollár et al [66], 2019 | PA; Xbox 360 | • Exergame was designed to improve postural control, gait mobility, gait stability, turning, and dynamic and static<br>• Exergame used the 3 visual feedback modules of the Xbox 360 core system, Kinect Adventures video game (Reflex Ridge, Space Pop, and Just Dance) |
| Tollár et al [67], 2019 | PA; Xbox 360 | • Exergame was designed to improve postural control, gait mobility, gait stability, turning, and balance<br>• Exergame used 3 Xbox 360 modules (Reflex Ridge, Space Pop, and Just Dance) |
| de Morais et al [63], 2020 | PA; Xbox 360 | • Xbox Kinect—"Your Shape Fitness evolved" (Zen-Develop it, Pump it, Wall Breacker, Kick it, Hurricane, and Stack in Up)<br>• The games are classified as easy, medium, or hard levels, and only the easy level was used |

| Authors, year | Type of serious game; device | Contents |
|---|---|---|
| Rica et al [61], 2020 | PA; Xbox 360 | • For Kinect-based exercise protocol, balance games were included<br>• Kinect Sports Ultimate Collection, Your Shape Fitness Evolved, Dance Central, and Nike + Kinect Training |
| Jahouh et al [68], 2021 | Both; Nintendo | • The intervention made up of different activities with the Nintendo Wii Fit video game console<br>• An aerobic-type game was used as a warm-up exercise<br>• The next game was played specifically to work on attention, concentration, and memory. In this game, a goalkeeper throws balls or bears from both the left and right sides. The participant was required to lean to either side to avoid all possible bears and head all possible balls; in other words, the participants had to swing on the same side of the ball or on the opposite side of the bear<br>• To end the session, the participants had to choose a game that they wanted to try or play |
| Kang et al [65], 2021 | CF; Oculus Rift CV1 and Oculus touch controllers | • Training was accompanied by game elements to increase the interest and motivation of the participants<br>• Games involving multidomain cognitive tasks to assess |
| Swinnen et al [71], 2021 | Both; Dividat Senso | • Dividat Senso consisted of a step training platform that was sensitive to pressure changes<br>• The sensors detected steps in 4 directions: left, right, top, and bottom<br>• The platform was connected via a USB cable to a computer and a frontal television screen on which the exergames were displayed<br>• Participants interacted with the game interface by pushing foot on 1 of the 4 different arrows<br>• The games trained cognitive abilities<br>• The device provided real-time visual, auditory, and somatosensory (vibrating platform) cues and feedback to enrich the game experience |

[a] PA: physical activity.

[b] CF: cognitive function.

**Table 3.** Characteristics of serious game interventions of the included studies (N=17).

| Authors, year | Type of serious game | Duration, frequency, time per session, dose | Control group | Interventionist |
|---|---|---|---|---|
| Rendon et al [60], 2012 | PA[a] | 6 weeks, 3 sessions per week, 35-45 minutes, 18 sessions | Usual care | Physical therapist |
| Schoene et al [56], 2015 | Both | 16 weeks, 3 sessions per week, 20 minutes, 48 sessions | Brochure | NR[b] |
| Choi et al [59], 2016 | PA | 2 weeks, 5 sessions per week, 60 minutes, 10 sessions | Exercise (conventional occupational therapy) | Occupational therapist |
| Levy et al [69], 2016 | PA | 12 weeks, 1 session per week, <40 minutes, 12 sessions | Usual care | NR |
| Nouchi et al [72], 2016 | CF[c] | 4 weeks, 5 sessions per week, 15 minutes, 20 sessions | Nonserious game (knowledge quiz) | NR |
| Anguera et al [57], 2017 | CF | 8 weeks, 5 sessions per week (biweekly), <20 minutes, 20 sessions | Problem-solving therapy | Interprofessional team (clinicians, care managers, and therapists) |
| Ferraz et al [64], 2018 | PA | 8 weeks, 3 sessions per week, 50 minutes, 24 sessions | Exercise (functional training); exercise (bicycle) | Physical therapist |
| Belchior et al [58], 2019 | CF | 12 weeks, 5 sessions per week, 60 minutes, 60 sessions | Cognitive training; usual care | NR |
| Smith et al [62], 2019 | CF | 12 months, NR, NR (600 minutes per 5-6 weeks), NR | Nonserious game (computerized crossword puzzles) | NR |
| Stanmore et al [70], 2019 | PA | 12 weeks, 3 sessions per week, NR, 36 sessions | Usual care | Physical therapist |
| Tollár et al [66], 2019 | PA | 5 weeks, 5 sessions per week, 60 minutes, 25 sessions | Exercise (stationary cycling); usual care | Physical therapist |
| Tollár et al [67], 2019 | PA | 5 weeks, 5 sessions per week, 60 minutes, 25 sessions | Exercise (stationary cycling); usual care | Physical therapist |
| de Morais et al [63], 2020 | PA | NR, NR, 30 minutes, 2 sessions | Watching a film | NR |
| Rica et al [61], 2020 | PA | 12 weeks, 3 sessions per week, 60 minutes, 36 sessions | Nonserious game (board games) | NR |
| Jahouh et al [68], 2021 | Both | 8 weeks, 2-3 sessions per week, 40-45 minutes, 20 sessions | Usual care | NR |
| Kang et al [65], 2021 | CF | 4 weeks, 2 sessions per week, 20-30 minutes, 8 sessions | Usual care | Neuropsychologist |
| Swinnen et al [71], 2021 | Both | 8 weeks, 3 sessions per week, 15 minutes, 24 sessions | Watching music videos | Physical therapist |

[a] PA: physical activity.

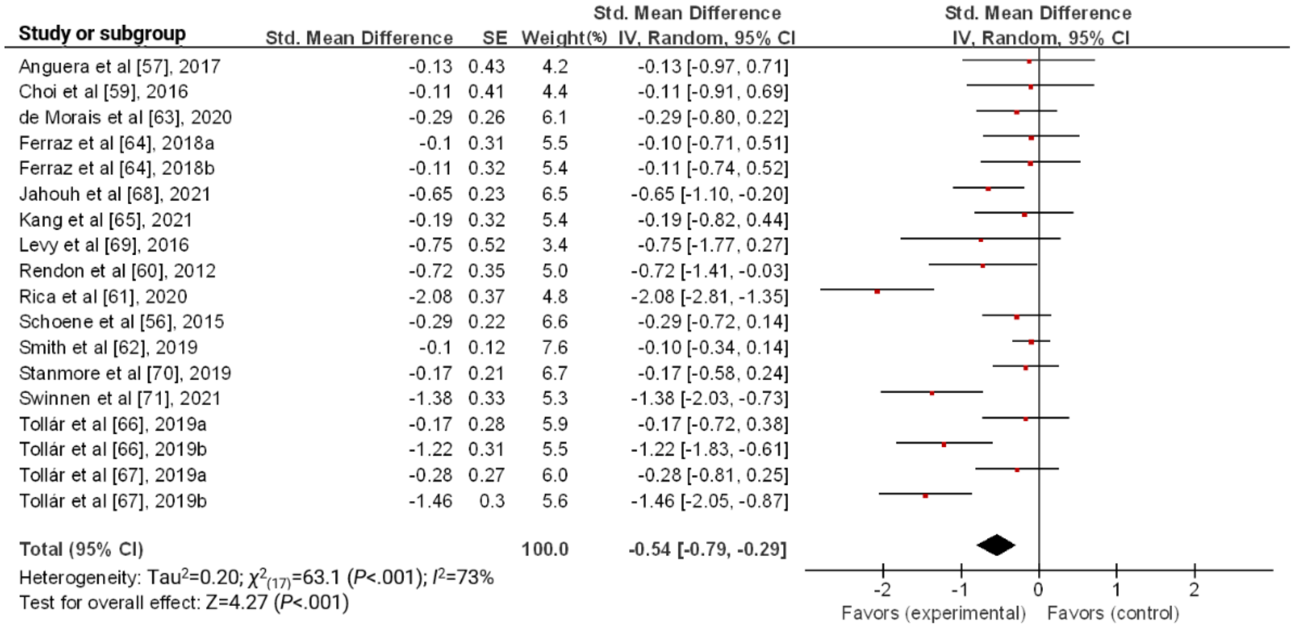
[b] NR: not reported.

[c] CF: cognitive function.

## Effects of Serious Games on Depression

Among the 17 studies included in the review, a meta-analysis was conducted on 15 (88%) studies, excluding 2 (12%) studies that did not provide raw data [58,72]. As 3 studies had 2 control groups each [64,66,67], we included 18 results in this meta-analysis. The pooled SMD between groups was −0.54 (95% CI −0.79 to −0.29; $P<.001$) with a medium effect size. These results indicate that serious games reduce depression in older adults. The heterogeneity of the meta-analysis was moderate to high across the studies ($I^2=73\%$; $P<.001$; Figure 2).

**Figure 2.** Forest plot for the effect of a serious game on depression.



| Study or subgroup | Std. Mean Difference | SE | Weight (%) | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|
| Anguera et al [57], 2017 | -0.13 | 0.43 | 4.2 | -0.13 [-0.97, 0.71] |
| Choi et al [59], 2016 | -0.11 | 0.41 | 4.4 | -0.11 [-0.91, 0.69] |
| de Morais et al [63], 2020 | -0.29 | 0.26 | 6.1 | -0.29 [-0.80, 0.22] |
| Ferraz et al [64], 2018a | -0.1 | 0.31 | 5.5 | -0.10 [-0.71, 0.51] |
| Ferraz et al [64], 2018b | -0.11 | 0.32 | 5.4 | -0.11 [-0.74, 0.52] |
| Jahouh et al [68], 2021 | -0.65 | 0.23 | 6.5 | -0.65 [-1.10, -0.20] |
| Kang et al [65], 2021 | -0.19 | 0.32 | 5.4 | -0.19 [-0.82, 0.44] |
| Levy et al [69], 2016 | -0.75 | 0.52 | 3.4 | -0.75 [-1.77, 0.27] |
| Rendon et al [60], 2012 | -0.72 | 0.35 | 5.0 | -0.72 [-1.41, -0.03] |
| Rica et al [61], 2020 | -2.08 | 0.37 | 4.8 | -2.08 [-2.81, -1.35] |
| Schoene et al [56], 2015 | -0.29 | 0.22 | 6.6 | -0.29 [-0.72, 0.14] |
| Smith et al [62], 2019 | -0.1 | 0.12 | 7.6 | -0.10 [-0.34, 0.14] |
| Stanmore et al [70], 2019 | -0.17 | 0.21 | 6.7 | -0.17 [-0.58, 0.24] |
| Swinnen et al [71], 2021 | -1.38 | 0.33 | 5.3 | -1.38 [-2.03, -0.73] |
| Tollár et al [66], 2019a | -0.17 | 0.28 | 5.9 | -0.17 [-0.72, 0.38] |
| Tollár et al [66], 2019b | -1.22 | 0.31 | 5.5 | -1.22 [-1.83, -0.61] |
| Tollár et al [67], 2019a | -0.28 | 0.27 | 6.0 | -0.28 [-0.81, 0.25] |
| Tollár et al [67], 2019b | -1.46 | 0.3 | 5.6 | -1.46 [-2.05, -0.87] |
| **Total (95% CI)** | | | **100.0** | **-0.54 [-0.79, -0.29]** |

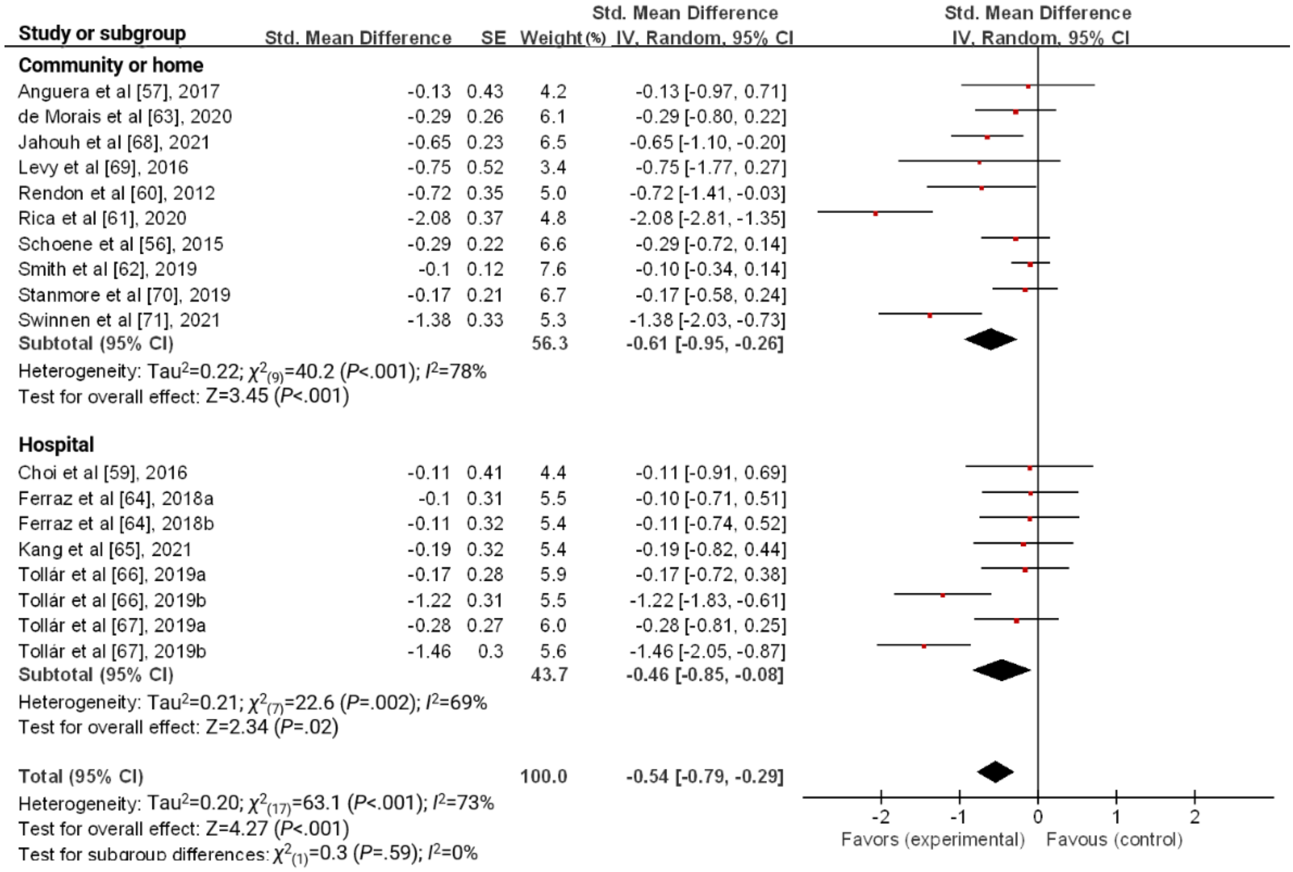
Heterogeneity: Tau$^2$=0.20; $\chi^2_{(17)}$=63.1 ($P$<.001); $I^2$=73%

Test for overall effect: Z=4.27 ($P$<.001)

## Subgroup Analysis

The results of the subgroup analysis of serious games for depression are shown in Figures 3-6. Regarding the setting, serious games had a more significant effect size in communities or homes (SMD −0.61, 95% CI −0.95 to −0.26; $P$<.001) than in hospitals (SMD −0.46, 95% CI −0.85 to −0.08; $P$=.02). However, the difference in the effect size between the groups was not statistically significant ($\chi^2_1$=0.3; $P$=.59; Figure 3).

**Figure 3.** Forest plot for the effect of a serious game on depression according to setting.



| Study or subgroup | Std. Mean Difference | SE | Weight (%) | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|
| **Community or home** | | | | |
| Anguera et al [57], 2017 | -0.13 | 0.43 | 4.2 | -0.13 [-0.97, 0.71] |
| de Morais et al [63], 2020 | -0.29 | 0.26 | 6.1 | -0.29 [-0.80, 0.22] |
| Jahouh et al [68], 2021 | -0.65 | 0.23 | 6.5 | -0.65 [-1.10, -0.20] |
| Levy et al [69], 2016 | -0.75 | 0.52 | 3.4 | -0.75 [-1.77, 0.27] |
| Rendon et al [60], 2012 | -0.72 | 0.35 | 5.0 | -0.72 [-1.41, -0.03] |
| Rica et al [61], 2020 | -2.08 | 0.37 | 4.8 | -2.08 [-2.81, -1.35] |
| Schoene et al [56], 2015 | -0.29 | 0.22 | 6.6 | -0.29 [-0.72, 0.14] |
| Smith et al [62], 2019 | -0.1 | 0.12 | 7.6 | -0.10 [-0.34, 0.14] |
| Stanmore et al [70], 2019 | -0.17 | 0.21 | 6.7 | -0.17 [-0.58, 0.24] |
| Swinnen et al [71], 2021 | -1.38 | 0.33 | 5.3 | -1.38 [-2.03, -0.73] |
| **Subtotal (95% CI)** | | | **56.3** | **-0.61 [-0.95, -0.26]** |
| Heterogeneity: Tau$^2$=0.22; $\chi^2_{(9)}$=40.2 ($P$<.001); $I^2$=78% | | | | |
| Test for overall effect: Z=3.45 ($P$<.001) | | | | |
| **Hospital** | | | | |
| Choi et al [59], 2016 | -0.11 | 0.41 | 4.4 | -0.11 [-0.91, 0.69] |
| Ferraz et al [64], 2018a | -0.1 | 0.31 | 5.5 | -0.10 [-0.71, 0.51] |
| Ferraz et al [64], 2018b | -0.11 | 0.32 | 5.4 | -0.11 [-0.74, 0.52] |
| Kang et al [65], 2021 | -0.19 | 0.32 | 5.4 | -0.19 [-0.82, 0.44] |
| Tollár et al [66], 2019a | -0.17 | 0.28 | 5.9 | -0.17 [-0.72, 0.38] |
| Tollár et al [66], 2019b | -1.22 | 0.31 | 5.5 | -1.22 [-1.83, -0.61] |
| Tollár et al [67], 2019a | -0.28 | 0.27 | 6.0 | -0.28 [-0.81, 0.25] |
| Tollár et al [67], 2019b | -1.46 | 0.3 | 5.6 | -1.46 [-2.05, -0.87] |
| **Subtotal (95% CI)** | | | **43.7** | **-0.46 [-0.85, -0.08]** |
| Heterogeneity: Tau$^2$=0.21; $\chi^2_{(7)}$=22.6 ($P$=.002); $I^2$=69% | | | | |
| Test for overall effect: Z=2.34 ($P$=.02) | | | | |
| **Total (95% CI)** | | | **100.0** | **-0.54 [-0.79, -0.29]** |

Heterogeneity: Tau$^2$=0.20; $\chi^2_{(17)}$=63.1 ($P$<.001); $I^2$=73%

Test for overall effect: Z=4.27 ($P$<.001)

Test for subgroup differences: $\chi^2_{(1)}$=0.3 ($P$=.59); $I^2$=0%

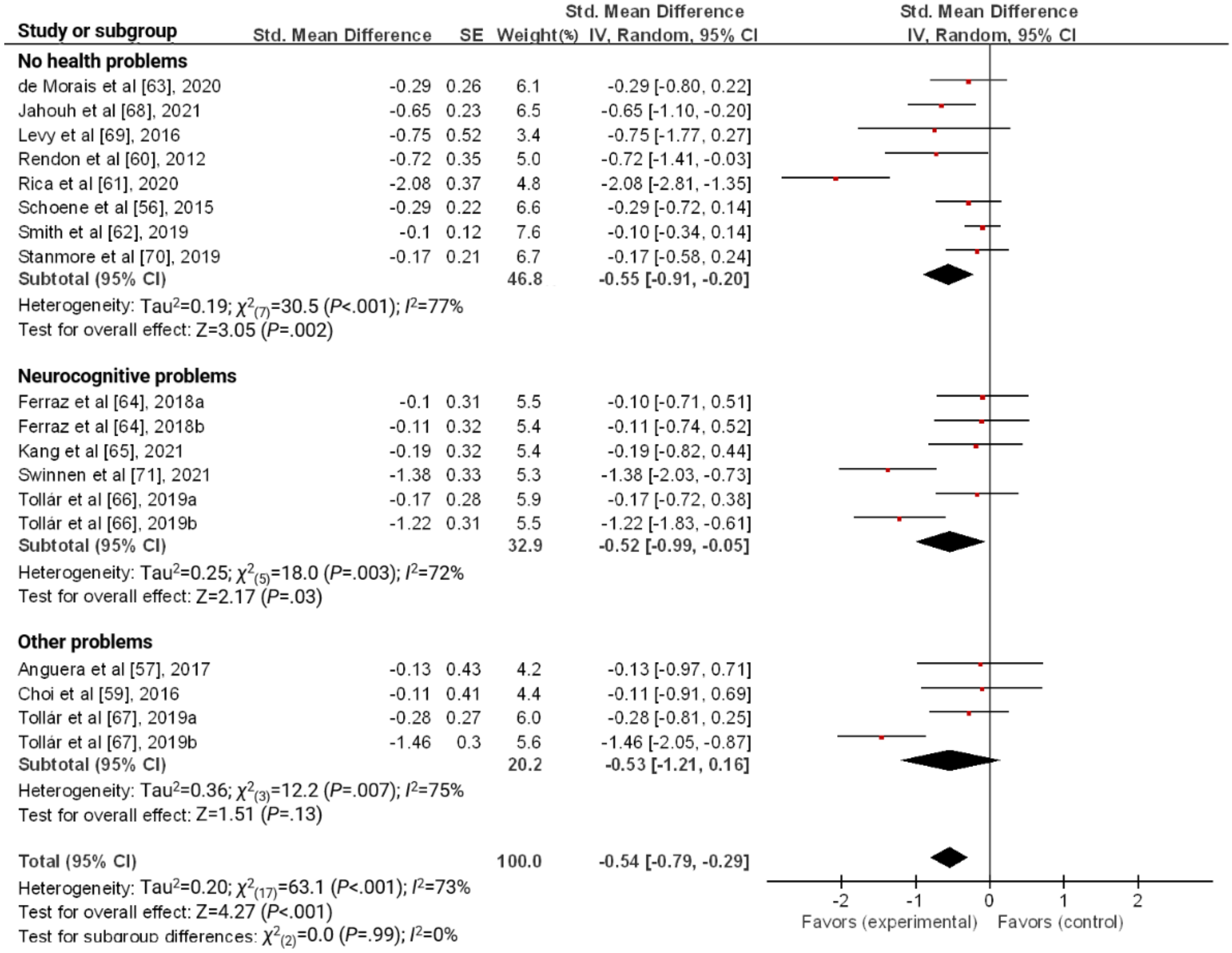
Regarding the characteristics of participants, the effect sizes of participants without health problems and with neurocognitive problems were −0.55 (95% CI −0.91 to −0.20; $P$=.002) and −0.52 (95% CI −0.99 to −0.05; $P$=.03), respectively, which significantly reduced depression (Figure 4).

A subgroup analysis of the type of serious game revealed that games for both had a significant effect on reducing depression in older adults (SMD −0.73, 95% CI −1.29 to −0.17; *P*=.01). In addition, games for PA significantly reduced depression (SMD −0.60, 95% CI −0.95 to −0.25; *P*<.001), whereas there was no significant effect of games on CF (Figure 5).

In the control group, serious games versus usual care had a significant effect on reducing depression (SMD −0.72, 95% CI −1.10 to −0.33; *P*<.001). However, subgroups of serious games versus other active comparators such as exercise, nonserious games, watching a film or music videos, and other programs presented no significant effect (Figure 6).

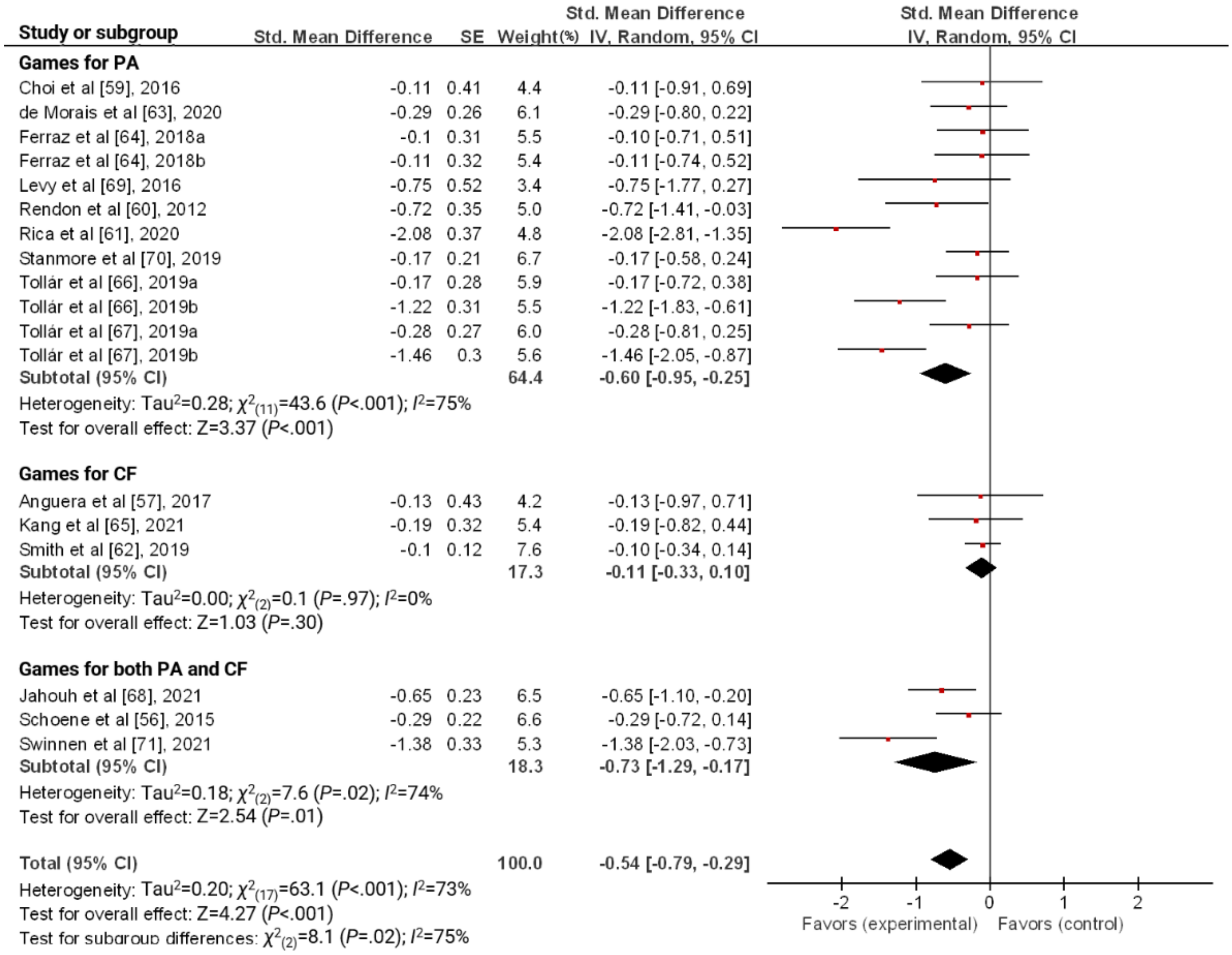
The results of the meta-regression indicated that no significant correlation existed between depression and the duration (*P*=.40) or dose of serious games (*P*=.43; Multimedia Appendix 3).

**Figure 4.** Forest plot for the effect of a serious game on depression according to the characteristics of participants.



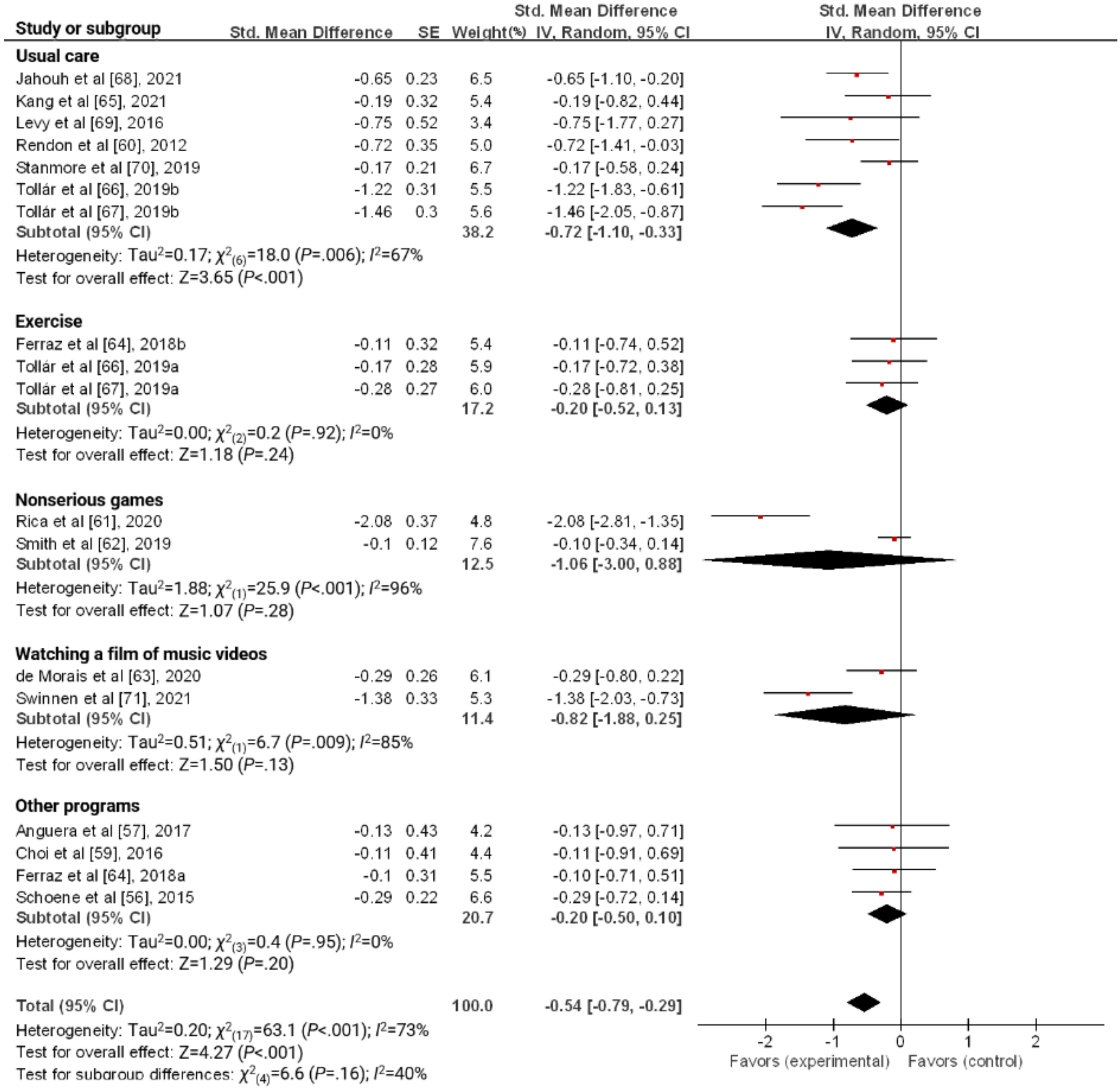
| Study or subgroup | Std. Mean Difference | SE | Weight(%) | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|
| **No health problems** | | | | |
| de Morais et al [63], 2020 | -0.29 | 0.26 | 6.1 | -0.29 [-0.80, 0.22] |
| Jahouh et al [68], 2021 | -0.65 | 0.23 | 6.5 | -0.65 [-1.10, -0.20] |
| Levy et al [69], 2016 | -0.75 | 0.52 | 3.4 | -0.75 [-1.77, 0.27] |
| Rendon et al [60], 2012 | -0.72 | 0.35 | 5.0 | -0.72 [-1.41, -0.03] |
| Rica et al [61], 2020 | -2.08 | 0.37 | 4.8 | -2.08 [-2.81, -1.35] |
| Schoene et al [56], 2015 | -0.29 | 0.22 | 6.6 | -0.29 [-0.72, 0.14] |
| Smith et al [62], 2019 | -0.1 | 0.12 | 7.6 | -0.10 [-0.34, 0.14] |
| Stanmore et al [70], 2019 | -0.17 | 0.21 | 6.7 | -0.17 [-0.58, 0.24] |
| **Subtotal (95% CI)** | | | **46.8** | **-0.55 [-0.91, -0.20]** |
| Heterogeneity: Tau$^2$=0.19; $\chi^2_{(7)}$=30.5 (*P*<.001); *I*$^2$=77% | | | | |
| Test for overall effect: Z=3.05 (*P*=.002) | | | | |
| **Neurocognitive problems** | | | | |
| Ferraz et al [64], 2018a | -0.1 | 0.31 | 5.5 | -0.10 [-0.71, 0.51] |
| Ferraz et al [64], 2018b | -0.11 | 0.32 | 5.4 | -0.11 [-0.74, 0.52] |
| Kang et al [65], 2021 | -0.19 | 0.32 | 5.4 | -0.19 [-0.82, 0.44] |
| Swinnen et al [71], 2021 | -1.38 | 0.33 | 5.3 | -1.38 [-2.03, -0.73] |
| Tollár et al [66], 2019a | -0.17 | 0.28 | 5.9 | -0.17 [-0.72, 0.38] |
| Tollár et al [66], 2019b | -1.22 | 0.31 | 5.5 | -1.22 [-1.83, -0.61] |
| **Subtotal (95% CI)** | | | **32.9** | **-0.52 [-0.99, -0.05]** |
| Heterogeneity: Tau$^2$=0.25; $\chi^2_{(5)}$=18.0 (*P*=.003); *I*$^2$=72% | | | | |
| Test for overall effect: Z=2.17 (*P*=.03) | | | | |
| **Other problems** | | | | |
| Anguera et al [57], 2017 | -0.13 | 0.43 | 4.2 | -0.13 [-0.97, 0.71] |
| Choi et al [59], 2016 | -0.11 | 0.41 | 4.4 | -0.11 [-0.91, 0.69] |
| Tollár et al [67], 2019a | -0.28 | 0.27 | 6.0 | -0.28 [-0.81, 0.25] |
| Tollár et al [67], 2019b | -1.46 | 0.3 | 5.6 | -1.46 [-2.05, -0.87] |
| **Subtotal (95% CI)** | | | **20.2** | **-0.53 [-1.21, 0.16]** |
| Heterogeneity: Tau$^2$=0.36; $\chi^2_{(3)}$=12.2 (*P*=.007); *I*$^2$=75% | | | | |
| Test for overall effect: Z=1.51 (*P*=.13) | | | | |
| **Total (95% CI)** | | | **100.0** | **-0.54 [-0.79, -0.29]** |
| Heterogeneity: Tau$^2$=0.20; $\chi^2_{(17)}$=63.1 (*P*<.001); *I*$^2$=73% | | | | |
| Test for overall effect: Z=4.27 (*P*<.001) | | | | |
| Test for subgroup differences: $\chi^2_{(2)}$=0.0 (*P*=.99); *I*$^2$=0% | | | | |

**Figure 5.** Forest plot for the effect of a serious game on depression according to the type of serious games. CF: cognitive function; PA: physical activity.



| Study or subgroup | Std. Mean Difference | SE | Weight(%) | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|
| **Games for PA** | | | | |
| Choi et al [59], 2016 | -0.11 | 0.41 | 4.4 | -0.11 [-0.91, 0.69] |
| de Morais et al [63], 2020 | -0.29 | 0.26 | 6.1 | -0.29 [-0.80, 0.22] |
| Ferraz et al [64], 2018a | -0.1 | 0.31 | 5.5 | -0.10 [-0.71, 0.51] |
| Ferraz et al [64], 2018b | -0.11 | 0.32 | 5.4 | -0.11 [-0.74, 0.52] |
| Levy et al [69], 2016 | -0.75 | 0.52 | 3.4 | -0.75 [-1.77, 0.27] |
| Rendon et al [60], 2012 | -0.72 | 0.35 | 5.0 | -0.72 [-1.41, -0.03] |
| Rica et al [61], 2020 | -2.08 | 0.37 | 4.8 | -2.08 [-2.81, -1.35] |
| Stanmore et al [70], 2019 | -0.17 | 0.21 | 6.7 | -0.17 [-0.58, 0.24] |
| Tollár et al [66], 2019a | -0.17 | 0.28 | 5.9 | -0.17 [-0.72, 0.38] |
| Tollár et al [66], 2019b | -1.22 | 0.31 | 5.5 | -1.22 [-1.83, -0.61] |
| Tollár et al [67], 2019a | -0.28 | 0.27 | 6.0 | -0.28 [-0.81, 0.25] |
| Tollár et al [67], 2019b | -1.46 | 0.3 | 5.6 | -1.46 [-2.05, -0.87] |
| **Subtotal (95% CI)** | | | **64.4** | **-0.60 [-0.95, -0.25]** |
| Heterogeneity: Tau$^2$=0.28; $\chi^2_{(11)}$=43.6 (*P*<.001); $I^2$=75% | | | | |
| Test for overall effect: Z=3.37 (*P*<.001) | | | | |
| **Games for CF** | | | | |
| Anguera et al [57], 2017 | -0.13 | 0.43 | 4.2 | -0.13 [-0.97, 0.71] |
| Kang et al [65], 2021 | -0.19 | 0.32 | 5.4 | -0.19 [-0.82, 0.44] |
| Smith et al [62], 2019 | -0.1 | 0.12 | 7.6 | -0.10 [-0.34, 0.14] |
| **Subtotal (95% CI)** | | | **17.3** | **-0.11 [-0.33, 0.10]** |
| Heterogeneity: Tau$^2$=0.00; $\chi^2_{(2)}$=0.1 (*P*=.97); $I^2$=0% | | | | |
| Test for overall effect: Z=1.03 (*P*=.30) | | | | |
| **Games for both PA and CF** | | | | |
| Jahouh et al [68], 2021 | -0.65 | 0.23 | 6.5 | -0.65 [-1.10, -0.20] |
| Schoene et al [56], 2015 | -0.29 | 0.22 | 6.6 | -0.29 [-0.72, 0.14] |
| Swinnen et al [71], 2021 | -1.38 | 0.33 | 5.3 | -1.38 [-2.03, -0.73] |
| **Subtotal (95% CI)** | | | **18.3** | **-0.73 [-1.29, -0.17]** |
| Heterogeneity: Tau$^2$=0.18; $\chi^2_{(2)}$=7.6 (*P*=.02); $I^2$=74% | | | | |
| Test for overall effect: Z=2.54 (*P*=.01) | | | | |
| **Total (95% CI)** | | | **100.0** | **-0.54 [-0.79, -0.29]** |
| Heterogeneity: Tau$^2$=0.20; $\chi^2_{(17)}$=63.1 (*P*<.001); $I^2$=73% | | | | |
| Test for overall effect: Z=4.27 (*P*<.001) | | | | |
| Test for subgroup differences: $\chi^2_{(2)}$=8.1 (*P*=.02); $I^2$=75% | | | | |

**Figure 6.** Forest plot for the effect of a serious game on depression according to the type of control group.



| Study or subgroup | Std. Mean Difference | SE | Weight(%) | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|
| **Usual care** | | | | |
| Jahouh et al [68], 2021 | -0.65 | 0.23 | 6.5 | -0.65 [-1.10, -0.20] |
| Kang et al [65], 2021 | -0.19 | 0.32 | 5.4 | -0.19 [-0.82, 0.44] |
| Levy et al [69], 2016 | -0.75 | 0.52 | 3.4 | -0.75 [-1.77, 0.27] |
| Rendon et al [60], 2012 | -0.72 | 0.35 | 5.0 | -0.72 [-1.41, -0.03] |
| Stanmore et al [70], 2019 | -0.17 | 0.21 | 6.7 | -0.17 [-0.58, 0.24] |
| Tollár et al [66], 2019b | -1.22 | 0.31 | 5.5 | -1.22 [-1.83, -0.61] |
| Tollár et al [67], 2019b | -1.46 | 0.3 | 5.6 | -1.46 [-2.05, -0.87] |
| Subtotal (95% CI) | | | 38.2 | -0.72 [-1.10, -0.33] |
| Heterogeneity: Tau$^2$=0.17; $\chi^2_{(6)}$=18.0 ($P$=.006); $I^2$=67% | | | | |
| Test for overall effect: Z=3.65 ($P$<.001) | | | | |
| **Exercise** | | | | |
| Ferraz et al [64], 2018b | -0.11 | 0.32 | 5.4 | -0.11 [-0.74, 0.52] |
| Tollár et al [66], 2019a | -0.17 | 0.28 | 5.9 | -0.17 [-0.72, 0.38] |
| Tollár et al [67], 2019a | -0.28 | 0.27 | 6.0 | -0.28 [-0.81, 0.25] |
| Subtotal (95% CI) | | | 17.2 | -0.20 [-0.52, 0.13] |
| Heterogeneity: Tau$^2$=0.00; $\chi^2_{(2)}$=0.2 ($P$=.92); $I^2$=0% | | | | |
| Test for overall effect: Z=1.18 ($P$=.24) | | | | |
| **Nonserious games** | | | | |
| Rica et al [61], 2020 | -2.08 | 0.37 | 4.8 | -2.08 [-2.81, -1.35] |
| Smith et al [62], 2019 | -0.1 | 0.12 | 7.6 | -0.10 [-0.34, 0.14] |
| Subtotal (95% CI) | | | 12.5 | -1.06 [-3.00, 0.88] |
| Heterogeneity: Tau$^2$=1.88; $\chi^2_{(1)}$=25.9 ($P$<.001); $I^2$=96% | | | | |
| Test for overall effect: Z=1.07 ($P$=.28) | | | | |
| **Watching a film of music videos** | | | | |
| de Morais et al [63], 2020 | -0.29 | 0.26 | 6.1 | -0.29 [-0.80, 0.22] |
| Swinnen et al [71], 2021 | -1.38 | 0.33 | 5.3 | -1.38 [-2.03, -0.73] |
| Subtotal (95% CI) | | | 11.4 | -0.82 [-1.88, 0.25] |
| Heterogeneity: Tau$^2$=0.51; $\chi^2_{(1)}$=6.7 ($P$=.009); $I^2$=85% | | | | |
| Test for overall effect: Z=1.50 ($P$=.13) | | | | |
| **Other programs** | | | | |
| Anguera et al [57], 2017 | -0.13 | 0.43 | 4.2 | -0.13 [-0.97, 0.71] |
| Choi et al [59], 2016 | -0.11 | 0.41 | 4.4 | -0.11 [-0.91, 0.69] |
| Ferraz et al [64], 2018a | -0.1 | 0.31 | 5.5 | -0.10 [-0.71, 0.51] |
| Schoene et al [56], 2015 | -0.29 | 0.22 | 6.6 | -0.29 [-0.72, 0.14] |
| Subtotal (95% CI) | | | 20.7 | -0.20 [-0.50, 0.10] |
| Heterogeneity: Tau$^2$=0.00; $\chi^2_{(3)}$=0.4 ($P$=.95); $I^2$=0% | | | | |
| Test for overall effect: Z=1.29 ($P$=.20) | | | | |
| Total (95% CI) | | | 100.0 | -0.54 [-0.79, -0.29] |
| Heterogeneity: Tau$^2$=0.20; $\chi^2_{(17)}$=63.1 ($P$<.001); $I^2$=73% | | | | |
| Test for overall effect: Z=4.27 ($P$<.001) | | | | |
| Test for subgroup differences: $\chi^2_{(4)}$=6.6 ($P$=.16); $I^2$=40% | | | | |

## Publication Bias

In this study, the publication bias was considered low. The funnel plot revealed a fairly symmetrical pattern (Multimedia Appendix 4). In addition, Egger test demonstrated that no publication bias was present in this meta-analysis ($P$=.27).

# *Discussion*

## Principal Findings

This study aimed to investigate the characteristics of serious games and their effects on depression in older adults. A total of 17 studies included in the systematic review have been conducted since 2012, and the number of studies has been steadily increasing. This study identified 3 types of serious games: games for PA, games for CF, and games for both PA and CF. Furthermore, it was demonstrated that the equipment and components of the game varied for each type of serious game. In addition, this study could provide substantial scientific evidence and produce high-quality findings as it included randomized controlled trials.

Older adults may face certain barriers to the use of digital interventions, such as physical changes because of aging and a lack of knowledge about technology [73]. In addition, studies included in this review reported that older adults who participated in serious games might experience difficulties with costs related to devices or programs [58], use of technology [70], and physical symptoms such as nausea, oculomotor dysfunction, and disorientation [65]. Despite these obstacles, the results of our meta-analysis indicate that serious games significantly reduce depression in older adults to a moderate effect size (SMD=−0.54). This aligns with the meta-analysis findings that showed reduced depression in young people [45]. In addition, the effect size of our findings was larger than that of studies targeting the general population [44,74]. Compared with the young or general population, serious games may effectively reduce depression in older adults, who face barriers in the application of digital interventions. As overcoming obstacles related to digital intervention may increase the interest and confidence of older adults [75], a serious game is considered effective and acceptable for them.

In this study, interventions of serious games were primarily conducted in places where older adults lived their daily lives, such as communities and homes, rather than in hospitals. Our findings showed that serious games played in communities or homes significantly reduced depression. Even without a supervised environment, such as a hospital, we found that the settings did not significantly affect older adults in applying serious games. This can be particularly convenient when applying the intervention to older adults with poor mobility or other accompanying diseases [47]. Therefore, a serious game can be applied regardless of the location once the appropriate environment or equipment is prepared.

In a subgroup analysis, serious games reduced depression in older adults without health problems or with neurocognitive problems. Appropriate physical abilities are required to perform serious games [76]; the intervention was effectively provided to older adults without health problems. Moreover, depression is considered an important health issue in older adults with neurocognitive problems such as Parkinson disease [77] and the predementia stage [78]. Accordingly, the findings of this study might be promising, as serious games may help reduce depression. However, the evidence may be relatively weak owing to the small number of studies. Further intervention studies are needed to confirm the relationship between the characteristics of older adults and the effectiveness of games.

Among the types of games, those that applied PA, including games for both accounted for approximately 70%. Our findings indicated that helping body movements directly by using various devices had a significant effect on reducing depression compared with games for CF. PA has been found to reduce depression through biological and psychosocial mechanisms [79]. In addition, a previous meta-analysis illustrated that exercise can significantly reduce depression in older adults [80]. Games promoting PA suggest the possibility of improving the quality of life and reducing depressive symptoms in older adults [61]. Therefore, PA should be considered as an essential component in the application of serious games to manage depression in older adults.

In the meta-analysis, we included studies that provided a control group with usual care as well as studies that provided other interventions considered active comparators. For studies comparing usual care groups, participation in serious games was found to have a significant impact on reducing depression. However, studies comparing active comparators, such as exercise, nonserious games, and watching videos, showed a reduction in depression, but this was not statistically significant. These findings are consistent with those of a meta-analysis that included active comparators [81]. The absolute and relative effects of the intervention can be interpreted according to the type of control group [82]. When the control group received usual care, the results indicated an absolute effect of the intervention. However, when the control group had active comparators, the results demonstrated a relative effect of the intervention. In this study, the absolute effect of participation in serious games was confirmed, but the relative effect was not. These findings indicate that participation in serious games has a unique and significant effect on older adults. Among the advantages of serious games, older adults can participate in such games regardless of the location [59], and it is generally easy to participate in such games [57]. In addition, participation in serious games motivates older adults and improves engagement [58,61], which has been found to increase adherence to interventions [57,63]. Participation in serious games for older adults may be viewed as an acceptable strategy to reduce depressive symptoms, as it has been confirmed that they have relatively high interest, satisfaction, and usability in serious games [59,65,70]. Therefore, we suggest that it is necessary to develop serious games that reflect helpful characteristics so that the relative effect as well as the absolute effect of interventions can be confirmed.

## Limitations

This study systematically reviewed the literature and analyzed the effectiveness of an overall serious game conducted to manage depression, without limiting the health characteristics of older adults. However, this study has a few limitations. First, some studies in which depression was not the primary outcome of the intervention were also selected because depression itself was of interest in this study. Generally, this selection may have a weak causal relationship between serious games and depression. In this review of 17 studies, 9 (53%) and 8 (47%) studies measured depression as a primary and secondary outcome, respectively. Depression had a significant reduction effect in 5 out of 9 (56%) studies measured as a primary outcome and 38% (3/8) of studies measured as a secondary outcome. The effects are greater when the intervention is performed for the main purpose of reducing depression. Therefore, further research is needed to clarify the relationship between depression as an outcome of intervention and its effectiveness. Second, methodological quality appraisals were performed using the JBI Critical Appraisal Checklist for the 17 studies included in this review. Only 6% (1/17) of high-quality studies had a low risk of bias, and approximately one-third (n=5, 29%) of the studies had a high risk of bias. Therefore, it can be considered that the evidence of the synthesized results is relatively low. Finally, this study included a variety of health characteristics, devices, and contents of serious games, which might lead to moderate to high heterogeneity ($I^2$=73%). Thus, it is necessary to pay attention to the interpretation of the results.

## Implications

Our findings may contribute to the understanding of the effects of serious games on reducing depression in older adults. The findings of this study also provide researchers and health care providers with several implications for managing depression in older adults. First, there is a need to apply serious games involving PA to manage depression in older adults. It may be beneficial to add various types of serious games to increase their effectiveness. Second, serious games should be developed and adapted to suit the various characteristics and needs of older adults. In addition, it is necessary to further explore devices, content, and duration that will be effective for older adults. Finally, a large-scale and more rigorously designed randomized controlled trial of serious games should be conducted to provide scientific evidence.

## Conclusions

The findings of this review and meta-analysis demonstrate that serious games are beneficial for reducing depression in older adults. Among the types of serious games, those that include PA significantly reduce depression. Regarding the setting, interventions conducted in the community, including homes, can alleviate depression in older adults. We also found that studies provided by nurses or multidisciplinary teams were limited; therefore, nurse researchers should conduct serious game interventions further. In addition, more high-quality randomized controlled trials are needed to establish substantial evidence of the effectiveness of serious games on depression in older adults.

## Acknowledgments

This research was supported by the Brain Korea 21 FOUR Project funded by the National Research Foundation of Korea, Yonsei University College of Nursing, and Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2020R1A6A1A03041989). This research was also supported by a 2020 grant from Yonsei University College of Nursing.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies and results of this study.
[DOCX File , 18 KB - jmir_v24i9e37753_app1.docx ]

Multimedia Appendix 2
Methodological appraisal of the included studies (N=17).
[DOCX File , 21 KB - jmir_v24i9e37753_app2.docx ]

Multimedia Appendix 3
Results of the meta-regression for subgroups of a serious game on depression.
[DOCX File , 267 KB - jmir_v24i9e37753_app3.docx ]

Multimedia Appendix 4
Funnel plot of this study.
[DOCX File , 106 KB - jmir_v24i9e37753_app4.docx ]

## References

1. Griffin L, Patterson D, Mason TM, Vonnes C. Improving care coordination for patients 65 and older. Geriatr Nurs 2021;42(2):610-612. [doi: 10.1016/j.gerinurse.2021.03.004] [Medline: 33744007]
2. Mental health of older adults. World Health Organization. 2017 Dec 12. URL: https://www.who.int/en/news-room/fact-sheets/detail/mental-health-of-older-adults [accessed 2022-01-11]
3. Kim HS, Kang JS. Effect of a group music intervention on cognitive function and mental health outcomes among nursing home residents: a randomized controlled pilot study. Geriatr Nurs 2021;42(3):650-656. [doi: 10.1016/j.gerinurse.2021.03.012] [Medline: 33823423]
4. Lee KC, Tang WK, Bressington D. The experience of mindful yoga for older adults with depression. J Psychiatr Ment Health Nurs 2019 Apr;26(3-4):87-100. [doi: 10.1111/jpm.12517] [Medline: 30873714]
5. Choi M, Kong S, Jung D. Computer and internet interventions for loneliness and depression in older adults: a meta-analysis. Healthc Inform Res 2012 Sep;18(3):191-198 [FREE Full text] [doi: 10.4258/hir.2012.18.3.191] [Medline: 23115742]
6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC, USA: American Psychiatric Publishing; May 27, 2013.
7. Egeljić-Mihailović N, Brkić-Jovanović N, Krstić T, Simin D, Milutinović D. Social participation and depressive symptoms among older adults during the Covid-19 pandemic in Serbia: a cross-sectional study. Geriatr Nurs 2022;44:8-14 [FREE Full text] [doi: 10.1016/j.gerinurse.2021.12.014] [Medline: 34999368]
8. Siegmund LA, Distelhorst KS, Bena JF, Morrison SL. Relationships between physical activity, social isolation, and depression among older adults during COVID-19: a path analysis. Geriatr Nurs 2021;42(5):1240-1244 [FREE Full text] [doi: 10.1016/j.gerinurse.2021.08.012] [Medline: 34507833]

9. Charlton RA, Lamar M, Ajilore O, Kumar A. Preliminary analysis of age of illness onset effects on symptom profiles in major depressive disorder. Int J Geriatr Psychiatry 2013 Nov;28(11):1166-1174 [FREE Full text] [doi: 10.1002/gps.3939] [Medline: 23420677]
10. Czaja SJ, Moxley JH, Rogers WA. Social support, isolation, loneliness, and health among older adults in the PRISM randomized controlled trial. Front Psychol 2021 Oct 05;12:728658 [FREE Full text] [doi: 10.3389/fpsyg.2021.728658] [Medline: 34675843]
11. Kim C, Chang EJ, Kim CY. Regional differences in the effects of social relations on depression among Korean elderly and the moderating effect of living alone. J Prev Med Public Health 2021 Nov;54(6):441-450 [FREE Full text] [doi: 10.3961/jpmph.21.337] [Medline: 34875827]
12. Liu Y, Gellatly J. Barriers and facilitators of engagement in psychological therapies among older adults with depression: a systematic review and thematic synthesis. J Psychiatr Ment Health Nurs 2021 Aug;28(4):509-520. [doi: 10.1111/jpm.12697] [Medline: 33043526]
13. Apóstolo J, Bobrowicz-Campos E, Rodrigues M, Castro I, Cardoso D. The effectiveness of non-pharmacological interventions in older adults with depressive disorders: a systematic review. Int J Nurs Stud 2016 Jun;58:59-70. [doi: 10.1016/j.ijnurstu.2016.02.006] [Medline: 27087298]
14. Irwin MR, Carrillo C, Sadeghi N, Bjurstrom MF, Breen EC, Olmstead R. Prevention of incident and recurrent major depression in older adults with insomnia: a randomized clinical trial. JAMA Psychiatry 2022 Jan 01;79(1):33-41 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.3422] [Medline: 34817561]
15. Kassel MT, Rhodes E, Insel PS, Woodworth K, Garrison-Diehn C, Satre DD, et al. Cognitive outcomes are differentially associated with depression severity trajectories during psychotherapy treatment for late life major depressive disorder. Int J Geriatr Psychiatry 2022 Aug;37(8):e5779. [doi: 10.1002/gps.5779] [Medline: 35822633]
16. Holland JC, Weiss Wiesel T, Nelson CJ, Roth AJ, Alici Y. Geriatric Psycho-Oncology: A Quick Reference on the Psychosocial Dimensions of Cancer Symptom Management. Oxford, UK: Oxford University Press; Mar 03, 2015.
17. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
18. Zhao P, Yoo I, Lancey R, Varghese E. Mobile applications for pain management: an app analysis for clinical usage. BMC Med Inform Decis Mak 2019 May 30;19(1):106 [FREE Full text] [doi: 10.1186/s12911-019-0827-7] [Medline: 31146739]
19. Harerimana B, Forchuk C, O'Regan T. The use of technology for mental healthcare delivery among older adults with depressive symptoms: a systematic literature review. Int J Ment Health Nurs 2019 Jun;28(3):657-670. [doi: 10.1111/inm.12571] [Medline: 30666762]
20. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
21. Bostrom J, Sweeney G, Whiteson J, Dodson JA. Mobile health and cardiac rehabilitation in older adults. Clin Cardiol 2020 Feb;43(2):118-126 [FREE Full text] [doi: 10.1002/clc.23306] [Medline: 31825132]
22. Zaman SB, Khan RK, Evans RG, Thrift AG, Maddison R, Islam SM. Exploring barriers to and enablers of the adoption of information and communication technology for the care of older adults with chronic diseases: scoping review. JMIR Aging 2022 Jan 07;5(1):e25251 [FREE Full text] [doi: 10.2196/25251] [Medline: 34994695]
23. Eichenberg C, Schott M, Sawyer A, Aumayr G, Plößnig M. Feasibility and conceptualization of an e-Mental health treatment for depression in older adults: mixed-methods study. JMIR Aging 2018 Oct 23;1(2):e10973 [FREE Full text] [doi: 10.2196/10973] [Medline: 31518235]
24. Giunti G, Baum A, Giunta D, Plazzotta F, Benitez S, Gómez A, et al. Serious games: a concise overview on what they are and their potential applications to healthcare. Stud Health Technol Inform 2015;216:386-390. [doi: 10.3233/978-1-61499-564-7-386] [Medline: 26262077]
25. Vieira C, Ferreira da Silva Pais-Vieira C, Novais J, Perrotta A. Serious game design and clinical improvement in physical rehabilitation: systematic review. JMIR Serious Games 2021 Sep 23;9(3):e20066 [FREE Full text] [doi: 10.2196/20066] [Medline: 34554102]
26. Gillois P, Di Marco L, Nicolàs JM, Moreau-Gaudry A, Ego A, David-Tchouda S, et al. Biostatistics disruptive acculturation through serious gaming: a new hope. Stud Health Technol Inform 2020 Jun 16;270:1215-1216. [doi: 10.3233/SHTI200369] [Medline: 32570586]
27. Vinolo Gil MJ, Gonzalez-Medina G, Lucena-Anton D, Perez-Cabezas V, Ruiz-Molinero MD, Martín-Valero R. Augmented reality in physical therapy: systematic review and meta-analysis. JMIR Serious Games 2021 Dec 15;9(4):e30985 [FREE Full text] [doi: 10.2196/30985] [Medline: 34914611]
28. Masurovsky A. Controlling for placebo effects in computerized cognitive training studies with healthy older adults from 2016-2018: systematic review. JMIR Serious Games 2020 Jun 26;8(2):e14030 [FREE Full text] [doi: 10.2196/14030] [Medline: 32589159]

29. Yan M, Yin H, Meng Q, Wang S, Ding Y, Li G, et al. A virtual supermarket program for the screening of mild cognitive impairment in older adults: diagnostic accuracy study. JMIR Serious Games 2021 Dec 03;9(4):e30919 [FREE Full text] [doi: 10.2196/30919] [Medline: 34870610]
30. Luo Y, Li M, Tang J, Ren J, Zheng Y, Yu X, et al. Design of a virtual reality interactive training system for public health emergency preparedness for major emerging infectious diseases: theory and framework. JMIR Serious Games 2021 Dec 14;9(4):e29956 [FREE Full text] [doi: 10.2196/29956] [Medline: 34904951]
31. Abraham O, Rosenberger C, Tierney K, Birstler J. Investigating the use of a serious game to improve opioid safety awareness among adolescents: quantitative study. JMIR Serious Games 2021 Dec 23;9(4):e33975 [FREE Full text] [doi: 10.2196/33975] [Medline: 34747712]
32. Schakel L, Veldhuijzen DS, van Middendorp H, Prins C, Drittij AM, Vrieling F, et al. An internet-based psychological intervention with a serious game to improve vitality, psychological and physical condition, and immune function in healthy male adults: randomized controlled trial. J Med Internet Res 2020 Jul 24;22(7):e14861 [FREE Full text] [doi: 10.2196/14861] [Medline: 32706667]
33. Corregidor-Sánchez AI, Segura-Fragoso A, Rodríguez-Hernández M, Criado-Alvarez JJ, González-Gonzalez J, Polonio-López B. Can exergames contribute to improving walking capacity in older adults? A systematic review and meta-analysis. Maturitas 2020 Feb;132:40-48. [doi: 10.1016/j.maturitas.2019.12.006] [Medline: 31883662]
34. Paliokas I, Kalamaras E, Votis K, Doumpoulakis S, Lakka E, Kotsani M, et al. Using a virtual reality serious game to assess the performance of older adults with frailty. Adv Exp Med Biol 2020;1196:127-139. [doi: 10.1007/978-3-030-32637-1_13] [Medline: 32468314]
35. Yang C, Han X, Jin M, Xu J, Wang Y, Zhang Y, et al. The effect of video game-based interventions on performance and cognitive function in older adults: Bayesian network meta-analysis. JMIR Serious Games 2021 Dec 30;9(4):e27058 [FREE Full text] [doi: 10.2196/27058] [Medline: 34967759]
36. Alhasan H, Alshehri MA, Wheeler PC, Fong DT. Effects of interactive videogames on postural control and risk of fall outcomes in frail and pre-frail older adults: a systematic review and meta-analysis. Games Health J 2021 Apr;10(2):83-94. [doi: 10.1089/g4h.2020.0009] [Medline: 33651955]
37. Williams T, Kennedy-Malone L, Thompson J, Monge EC. The effect of an exergame on physical activity among older adults residing in a long-term care facility: a pilot study. Geriatr Nurs 2022;44:48-53. [doi: 10.1016/j.gerinurse.2022.01.001] [Medline: 35065365]
38. Zhu K, Zhang Q, He B, Huang M, Lin R, Li H. Immersive virtual reality-based cognitive intervention for the improvement of cognitive function, depression, and perceived stress in older adults with mild cognitive impairment and mild dementia: pilot pre-post study. JMIR Serious Games 2022 Feb 21;10(1):e32117 [FREE Full text] [doi: 10.2196/32117] [Medline: 35188466]
39. Beltran-Alacreu H, Navarro-Fernández G, Godia-Lledó D, Graell-Pasarón L, Ramos-González Á, Raya R, et al. A serious game for performing task-oriented cervical exercises among older adult patients with chronic neck pain: development, suitability, and crossover pilot study. JMIR Serious Games 2022 Feb 01;10(1):e31404 [FREE Full text] [doi: 10.2196/31404] [Medline: 35103608]
40. Merriman NA, Roudaia E, Ondřej J, Romagnoli M, Orvieto I, O'Sullivan C, et al. "CityQuest," a custom-designed serious game, enhances spatial memory performance in older adults. Front Aging Neurosci 2022 Mar 08;14:806418 [FREE Full text] [doi: 10.3389/fnagi.2022.806418] [Medline: 35356302]
41. Martinez K, Menéndez-Menéndez MI, Bustillo A. Awareness, prevention, detection, and therapy applications for depression and anxiety in serious games for children and adolescents: systematic review. JMIR Serious Games 2021 Dec 16;9(4):e30482 [FREE Full text] [doi: 10.2196/30482] [Medline: 34927589]
42. Novo A, Fonsêca J, Barroso B, Guimarães M, Louro A, Fernandes H, et al. Virtual reality rehabilitation's impact on negative symptoms and psychosocial rehabilitation in schizophrenia spectrum disorder: a systematic review. Healthcare (Basel) 2021 Oct 23;9(11):1429 [FREE Full text] [doi: 10.3390/healthcare9111429] [Medline: 34828476]
43. Siriaraya P, Visch V, Boffo M, Spijkerman R, Wiers R, Korrelboom K, et al. Game design in mental health care: case study-based framework for integrating game design into therapeutic content. JMIR Serious Games 2021 Dec 01;9(4):e27953 [FREE Full text] [doi: 10.2196/27953] [Medline: 34855611]
44. Li J, Theng YL, Foo S. Effect of exergames on depression: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2016 Jan;19(1):34-42. [doi: 10.1089/cyber.2015.0366] [Medline: 26716638]
45. Townsend C, Humpston C, Rogers J, Goodyear V, Lavis A, Michail M. The effectiveness of gaming interventions for depression and anxiety in young people: systematic review and meta-analysis. BJPsych Open 2022 Jan 07;8(1):e25 [FREE Full text] [doi: 10.1192/bjo.2021.1078] [Medline: 34991767]
46. Cacciata M, Stromberg A, Lee JA, Sorkin D, Lombardo D, Clancy S, et al. Effect of exergaming on health-related quality of life in older adults: a systematic review. Int J Nurs Stud 2019 May;93:30-40 [FREE Full text] [doi: 10.1016/j.ijnurstu.2019.01.010] [Medline: 30861452]
47. Chao YY, Scherer YK, Montgomery CA. Effects of using Nintendo Wii™ exergames in older adults: a review of the literature. J Aging Health 2015 Apr;27(3):379-402. [doi: 10.1177/0898264314551171] [Medline: 25245519]

48. Kappen DL, Mirza-Babaei P, Nacke LE. Older adults’ physical activity and exergames: a systematic review. Int J Human Comput Interact 2019;35(2):140-167. [doi: 10.1080/10447318.2018.1441253]
49. 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]
50. Aromataris E, Munn Z. JBI Manual for Evidence Synthesis. Joanna Briggs Institute. 2020. URL: https://synthesismanual.jbi.global [accessed 2021-07-06]
51. Tiruneh GT, Yakob B, Ayele WM, Yigzaw M, Roro MA, Medhanyi AA, et al. Effect of community-based distribution of misoprostol on facility delivery: a scoping review. BMC Pregnancy Childbirth 2019 Nov 06;19(1):404 [FREE Full text] [doi: 10.1186/s12884-019-2539-5] [Medline: 31694580]
52. Higgins JP, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.3. London, UK: The Cochrane Collaboration; 2022.
53. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
54. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ, USA: Lawrence Erlbaum Associates; 1988.
55. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
56. Schoene D, Valenzuela T, Toson B, Delbaere K, Severino C, Garcia J, et al. Interactive cognitive-motor step training improves cognitive risk factors of falling in older adults - a randomized controlled trial. PLoS One 2015 Dec 16;10(12):e0145161 [FREE Full text] [doi: 10.1371/journal.pone.0145161] [Medline: 26673919]
57. Anguera JA, Gunning FM, Areán PA. Improving late life depression and cognitive control through the use of therapeutic video game technology: a proof-of-concept randomized trial. Depress Anxiety 2017 Jun;34(6):508-517 [FREE Full text] [doi: 10.1002/da.22588] [Medline: 28052513]
58. Belchior P, Yam A, Thomas KR, Bavelier D, Ball KK, Mann WC, et al. Computer and videogame interventions for older adults' cognitive and everyday functioning. Games Health J 2019 Apr;8(2):129-143 [FREE Full text] [doi: 10.1089/g4h.2017.0092] [Medline: 30273002]
59. Choi YH, Ku J, Lim H, Kim YH, Paik NJ. Mobile game-based virtual reality rehabilitation program for upper limb dysfunction after ischemic stroke. Restor Neurol Neurosci 2016 May 02;34(3):455-463. [doi: 10.3233/RNN-150626] [Medline: 27163250]
60. Rendon AA, Lohman EB, Thorpe D, Johnson EG, Medina E, Bradley B. The effect of virtual reality gaming on dynamic balance in older adults. Age Ageing 2012 Jul;41(4):549-552. [doi: 10.1093/ageing/afs053] [Medline: 22672915]
61. Rica RL, Shimojo GL, Gomes MC, Alonso AC, Pitta RM, Santa-Rosa FA, et al. Effects of a Kinect-based physical training program on body composition, functional fitness and depression in institutionalized older adults. Geriatr Gerontol Int 2020 Mar;20(3):195-200. [doi: 10.1111/ggi.13857] [Medline: 31923924]
62. Smith M, Jones MP, Dotson MM, Wolinsky FD. Speed of processing training and depression in assisted and independent living: a randomized controlled trial. PLoS One 2019 Oct 17;14(10):e0223841 [FREE Full text] [doi: 10.1371/journal.pone.0223841] [Medline: 31622386]
63. de Morais MA, de Lima BE, Bandeira Santos LC. Acute effect of Xbox exercise on mood states in older adults. Act Adapt Aging 2020;44(2):146-156. [doi: 10.1080/01924788.2019.1618689]
64. Ferraz DD, Trippo KV, Duarte GP, Neto MG, Bernardes Santos KO, Filho JO. The effects of functional training, bicycle exercise, and exergaming on walking capacity of elderly patients with Parkinson disease: a pilot randomized controlled single-blinded trial. Arch Phys Med Rehabil 2018 May;99(5):826-833. [doi: 10.1016/j.apmr.2017.12.014] [Medline: 29337023]
65. Kang JM, Kim N, Lee SY, Woo SK, Park G, Yeon BK, et al. Effect of cognitive training in fully immersive virtual reality on visuospatial function and frontal-occipital functional connectivity in predementia: randomized controlled trial. J Med Internet Res 2021 May 06;23(5):e24526 [FREE Full text] [doi: 10.2196/24526] [Medline: 33955835]
66. Tollár J, Nagy F, Hortobágyi T. Vastly different exercise programs similarly improve Parkinsonian symptoms: a randomized clinical trial. Gerontology 2019;65(2):120-127. [doi: 10.1159/000493127] [Medline: 30368495]
67. Tollár J, Nagy F, Moizs M, Tóth BE, Sanders LM, Hortobágyi T. Diverse exercises similarly reduce older adults' mobility limitations. Med Sci Sports Exerc 2019 Sep;51(9):1809-1816. [doi: 10.1249/MSS.0000000000002001] [Medline: 30973482]
68. Jahouh M, González-Bernal JJ, González-Santos J, Fernández-Lázaro D, Soto-Cámara R, Mielgo-Ayuso J. Impact of an intervention with Wii video games on the autonomy of activities of daily living and psychological-cognitive components in the institutionalized elderly. Int J Environ Res Public Health 2021 Feb 07;18(4):1570 [FREE Full text] [doi: 10.3390/ijerph18041570] [Medline: 33562249]

69. Levy F, Leboucher P, Rautureau G, Komano O, Millet B, Jouvent R. Fear of falling: efficacy of virtual reality associated with serious games in elderly people. Neuropsychiatr Dis Treat 2016 Apr 15;12:877-881 [FREE Full text] [doi: 10.2147/NDT.S97809] [Medline: 27143889]
70. Stanmore EK, Mavroeidi A, de Jong LD, Skelton DA, Sutton CJ, Benedetto V, et al. The effectiveness and cost-effectiveness of strength and balance Exergames to reduce falls risk for people aged 55 years and older in UK assisted living facilities: a multi-centre, cluster randomised controlled trial. BMC Med 2019 Feb 28;17(1):49 [FREE Full text] [doi: 10.1186/s12916-019-1278-9] [Medline: 30813926]
71. Swinnen N, Vandenbulcke M, de Bruin ED, Akkerman R, Stubbs B, Firth J, et al. The efficacy of exergaming in people with major neurocognitive disorder residing in long-term care facilities: a pilot randomized controlled trial. Alzheimers Res Ther 2021 Mar 30;13(1):70 [FREE Full text] [doi: 10.1186/s13195-021-00806-7] [Medline: 33785077]
72. Nouchi R, Saito T, Nouchi H, Kawashima R. Small acute benefits of 4 weeks processing speed training games on processing speed and inhibition performance and depressive mood in the healthy elderly people: evidence from a randomized control trial. Front Aging Neurosci 2016 Dec 23;8:302 [FREE Full text] [doi: 10.3389/fnagi.2016.00302] [Medline: 28066229]
73. Gitlow L. Technology use by older adults and barriers to using technology. Phys Occup Ther Geriatr 2014 Aug 12;32(3):271-280. [doi: 10.3109/02703181.2014.946640]
74. Li J, Theng YL, Foo S. Game-based digital interventions for depression therapy: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):519-527 [FREE Full text] [doi: 10.1089/cyber.2013.0481] [Medline: 24810933]
75. Hill R, Betts LR, Gardner SE. Older adults' experiences and perceptions of digital technology: (dis)empowerment, wellbeing, and inclusion. Comput Human Behav 2015 Jul;48:415-423. [doi: 10.1016/j.chb.2015.01.062]
76. de Boissieu P, Denormandie P, Armaingaud D, Sanchez S, Jeandel C. Exergames and elderly: a non-systematic review of the literature. Eur Geriatr Med 2017 Apr;8(2):111-116. [doi: 10.1016/j.eurger.2017.02.003]
77. Ray S, Agarwal P. Depression and anxiety in Parkinson disease. Clin Geriatr Med 2020 Feb;36(1):93-104. [doi: 10.1016/j.cger.2019.09.012] [Medline: 31733705]
78. Ma L. Depression, anxiety, and apathy in mild cognitive impairment: current perspectives. Front Aging Neurosci 2020 Jan 30;12:9 [FREE Full text] [doi: 10.3389/fnagi.2020.00009] [Medline: 32082139]
79. Kandola A, Ashdown-Franks G, Hendrikse J, Sabiston CM, Stubbs B. Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity. Neurosci Biobehav Rev 2019 Dec;107:525-539. [doi: 10.1016/j.neubiorev.2019.09.040] [Medline: 31586447]
80. Schuch FB, Vancampfort D, Rosenbaum S, Richards J, Ward PB, Veronese N, et al. Exercise for depression in older adults: a meta-analysis of randomized controlled trials adjusting for publication bias. Braz J Psychiatry 2016;38(3):247-254 [FREE Full text] [doi: 10.1590/1516-4446-2016-1915] [Medline: 27611903]
81. Zhang Q, Fu Y, Lu Y, Zhang Y, Huang Q, Yang Y, et al. Impact of virtual reality-based therapies on cognition and mental health of stroke patients: systematic review and meta-analysis. J Med Internet Res 2021 Nov 17;23(11):e31007 [FREE Full text] [doi: 10.2196/31007] [Medline: 34787571]
82. Karlsson P, Bergmark A. Compared with what? An analysis of control-group types in Cochrane and Campbell reviews of psychosocial treatment efficacy with substance use disorders. Addiction 2015 Mar;110(3):420-428 [FREE Full text] [doi: 10.1111/add.12799] [Medline: 25393504]

## Abbreviations

**CF:** cognitive function
**PA:** physical activity
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**SMD:** standardized mean difference

*Edited by R Kukafka; submitted 07.03.22; peer-reviewed by YY Chao, D Sorkin; comments to author 28.03.22; revised version received 11.05.22; accepted 29.07.22; published 06.09.22.*

*Please cite as:*
*Kim Y, Hong S, Choi M*
*Effects of Serious Games on Depression in Older Adults: Systematic Review and Meta-analysis of Randomized Controlled Trials*
*J Med Internet Res 2022;24(9):e37753*
*URL: https://www.jmir.org/2022/9/e37753*
*doi:10.2196/37753*
*PMID:36066964*

Review

# Digital Health Interventions for Musculoskeletal Pain Conditions: Systematic Review and Meta-analysis of Randomized Controlled Trials

Pim Peter Valentijn[1,2], MSc, PhD; Liza Tymchenko[1], MSc; Teddy Jacobson[3], BSc; Jakob Kromann[3], BSc; Claus W Biermann[3], MSc, PhD; Mohamed Atef AlMoslemany[4]; Rosa Ymkje Arends[1,5], MSc, PhD

[1]Essenburgh Research & Consultancy, Essenburgh Group, Harderwijk, Netherlands

[2]Department of Health Services Research, School for Public Health and Primary Care, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[3]Nordic Netcare, Copenhagen, Denmark

[4]Faculty of Medicine, Menoufia University, Menoufia, Egypt

[5]University of Applied Sciences Utrecht, Utrecht, Netherlands

**Corresponding Author:**
Pim Peter Valentijn, MSc, PhD
Essenburgh Research & Consultancy
Essenburgh Group
Ceintuurbaan 2-40
Harderwijk, 3847 LG
Netherlands
Phone: 31 0341 217101
Email: valentijn@essenburgh.com

## Abstract

**Background:** Digital health solutions can provide populations with musculoskeletal pain with high-reach, low-cost, easily accessible, and scalable patient education and self-management interventions that meet the time and resource restrictions.

**Objective:** The main objective of this study was to determine the effectiveness of digital health interventions for people with musculoskeletal pain conditions (ie, low back pain, neck pain, shoulder pain, knee pain, elbow pain, ankle pain, and whiplash).

**Methods:** A systematic review and meta-analysis was conducted. We searched PubMed and Cochrane Central Register of Controlled Trials (from 1974 to August 2021) and selected randomized controlled trials of digital health interventions in the target population of patients with musculoskeletal pain with a minimum follow-up of 1 month. A total of 2 researchers independently screened and extracted the data.

**Results:** A total of 56 eligible studies were included covering 9359 participants, with a mean follow-up of 25 (SD 15.48) weeks. In moderate-quality evidence, digital health interventions had a small effect on pain (standardized mean difference [SMD] 0.19, 95% CI 0.06-0.32), disability (SMD 0.14, 95% CI 0.03-0.25), quality of life (SMD 0.22, 95% CI 0.07-0.36), emotional functioning (SMD 0.24, 95% CI 0.12-0.35), and self-management (SMD 0.14, 95% CI 0.05-0.24).

**Conclusions:** Moderate-quality evidence supports the conclusion that digital health interventions are effective in reducing pain and improving functioning and self-management of musculoskeletal pain conditions. Low-quality evidence indicates that digital health interventions can improve the quality of life and global treatment. Little research has been conducted on the influence of digital health on expenses, knowledge, overall improvement, range of motion, muscle strength, and implementation fidelity.

**Trial Registration:** PROSPERO CRD42022307504; https://tinyurl.com/2cd25hus

***(J Med Internet Res 2022;24(9):e37869)*** doi:10.2196/37869

**KEYWORDS**

eHealth; models of care; mobile health; mHealth; digital health; pain; telehealth; telemedicine; disability; function; quality of life; mobile phone

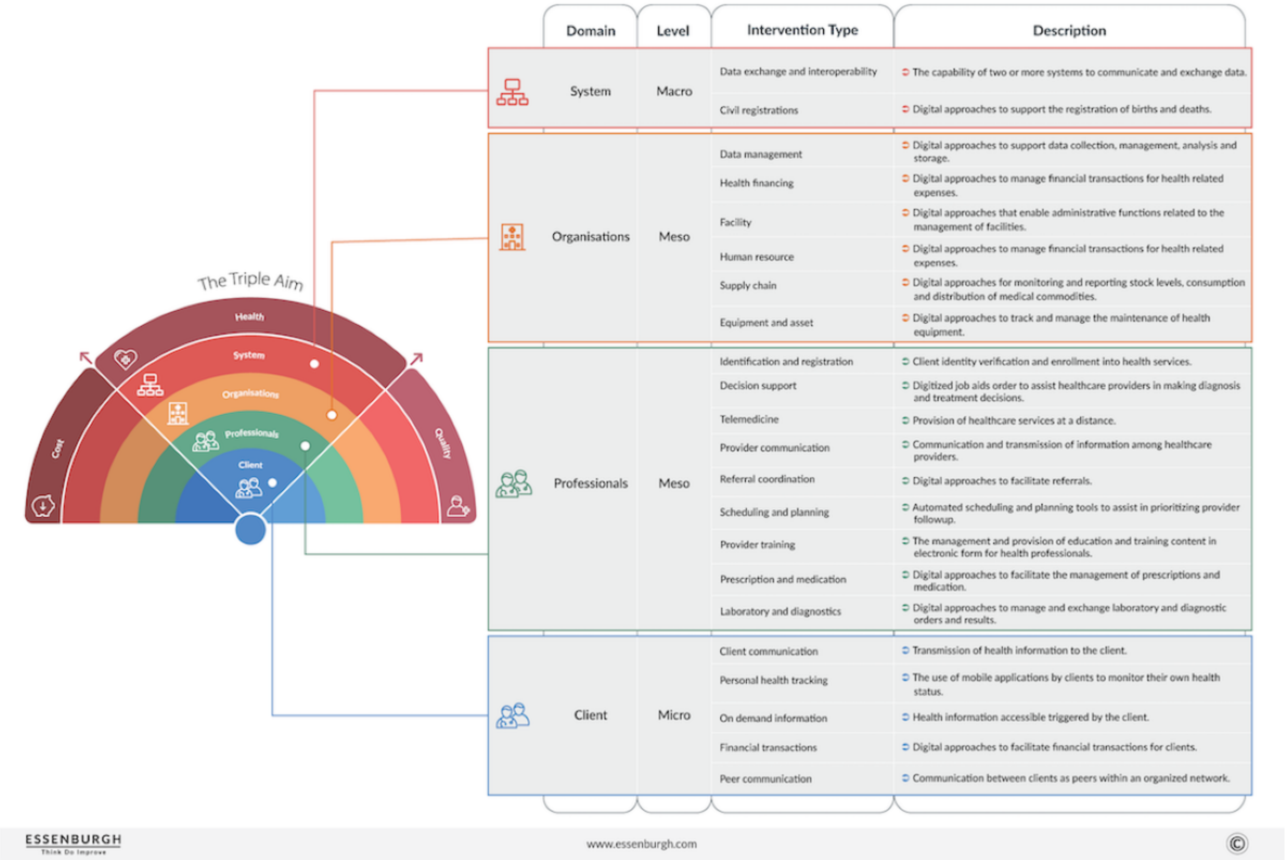
## Introduction

Musculoskeletal conditions are considered the leading cause of global morbidity and have substantial individual, societal, and economic implications [1]. Musculoskeletal conditions account for one-fifth of the world's total number of years lived with disability [1]. The burden of musculoskeletal conditions is predicted to increase dramatically in the coming years because of the aging population in Western countries. Musculoskeletal conditions include a broad range of health conditions affecting the bones, joints, muscles, and spine, as well as rare autoimmune conditions. Common symptoms include pain, stiffness, and loss of mobility and dexterity, which often interfere with people's ability to perform daily activities. In the global research on the burden of disease, low back and neck pain were responsible for 70% of impairments [2]. The management of musculoskeletal pain conditions requires an evidence-informed innovative care model that stimulates self-management, including daily activities, self-care, patient-professional collaboration, and a collaborative practice model [3].

For musculoskeletal pain conditions, there has been increasing interest in integrating digital health interventions to accomplish the triple aim of better health outcomes, better patient experiences, and smarter use of health service resources. Various studies have found moderate-quality evidence that digital health interventions have a positive clinical benefit in the management of musculoskeletal conditions leading to pain and functional disability [4-7]. However, owing to differences in content, duration, and delivery, it is difficult to draw strong conclusions about the effectiveness of digital health interventions. Hence, little is known about which type or combination of digital health solutions is superior [5,8-10]. This lack of information serves as a barrier to identifying key characteristics aligned with effective and ineffective digital health solutions and their wider implementation. Recently, the World Health Organization (WHO) published a taxonomy for the standardization of various digital health interventions and vocabulary [11]. Although taxonomy is a useful tool to differentiate between the different types of digital interventions, it cannot distinguish between the micro, meso, and macro factors that influence digital health innovation and implementation [12]. This calls for a broad overview of the evidence by outlining digital health solutions at the patient, professional, provider, and system levels, as described by the Rainbow Model (Figure 1) [13]. It is important to identify the most effective type of digital health intervention and, in turn, the most efficacious combination of components (eg, patient, provider, organizational, and system level) for clinical and managerial responses to the evidence, as well as for policy decision-making.

Following the Rainbow Model of Integrated Care (RMIC) and WHO digital health taxonomy, we comprehensively analyzed the effectiveness of digital health interventions for musculoskeletal pain conditions in published randomized controlled trials (RCTs) and assessed the extent to which differences in outcomes may be explained by the different types of interventions.

**Figure 1.** Rainbow model for digital health interventions [14].



| Domain | Level | Intervention Type | Description |
|---|---|---|---|
| System | Macro | Data exchange and interoperability | The capability of two or more systems to communicate and exchange data. |
| | | Civil registrations | Digital approaches to support the registration of births and deaths. |
| Organisations | Meso | Data management | Digital approaches to support data collection, management, analysis and storage. |
| | | Health financing | Digital approaches to manage financial transactions for health related expenses. |
| | | Facility | Digital approaches that enable administrative functions related to the management of facilities. |
| | | Human resource | Digital approaches to manage financial transactions for health related expenses. |
| | | Supply chain | Digital approaches for monitoring and reporting stock levels, consumption and distribution of medical commodities. |
| | | Equipment and asset | Digital approaches to track and manage the maintenance of health equipment. |
| Professionals | Meso | Identification and registration | Client identity verification and enrollment into health services. |
| | | Decision support | Digitized job aids order to assist healthcare providers in making diagnosis and treatment decisions. |
| | | Telemedicine | Provision of healthcare services at a distance. |
| | | Provider communication | Communication and transmission of information among healthcare providers. |
| | | Referral coordination | Digital approaches to facilitate referrals. |
| | | Scheduling and planning | Automated scheduling and planning tools to assist in prioritizing provider followup. |
| | | Provider training | The management and provision of education and training content in electronic form for health professionals. |
| | | Prescription and medication | Digital approaches to facilitate the management of prescriptions and medication. |
| | | Laboratory and diagnostics | Digital approaches to manage and exchange laboratory and diagnostic orders and results. |
| Client | Micro | Client communication | Transmission of health information to the client. |
| | | Personal health tracking | The use of mobile applications by clients to monitor their own health status. |
| | | On demand information | Health information accessible triggered by the client. |
| | | Financial transactions | Digital approaches to facilitate financial transactions for clients. |
| | | Peer communication | Communication between clients as peers within an organized network. |

## Methods

A systematic review was conducted according to a protocol registered on PROSPERO (International Prospective Register of Systematic Reviews; registration number 307504) and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15].

### Literature Search

We searched the electronic databases PubMed and the Cochrane Library using musculoskeletal pain condition–specific and digital health–specific text words and Medical Subject Headings (Tables S1 and S2 in Multimedia Appendix 1 [16-71]) from their inception to August 2021.

### Study Selection

#### Overview

A total of 2 researchers (LT and MAA) worked separately on study selection, eligibility criteria evaluation, risk-of-bias assessment, and data extraction, and disagreements were resolved through iteration and discussion. If this failed, a third arbitrary resolution was performed by a third author (PPV). Studies were considered eligible if they were RCTs with a follow-up of ≥1 month; included participants aged >18 years with a musculoskeletal (chronic) pain condition (ie, low back pain, neck pain, shoulder pain, knee pain, elbow pain, ankle pain, or whiplash); and comprised an evaluation of a digital health intervention in the clinical, professional, organizational, or system domains of the RMIC [13]. Each intervention had to describe ≥1 digital health service according to the description of the WHO digital health taxonomy [11] (Table S3 in Multimedia Appendix 1 [16-71]). Non-English studies were excluded from this review.

#### Data Extraction and Risk-of-Bias Assessment

For each included study, 2 researchers (LT and MAA) independently extracted the data using a standardized data extraction form. Any inconsistency was resolved through iteration and discussion. When the required data were not reported in the article, the researchers contacted the authors for the missing information. If the required data could not be provided, the study was included only for qualitative review. The following methodological risks of bias were assessed for each selected study: sequence generation; allocation concealment; blinding of outcome assessors, care providers, and participants; completeness of outcome data; intention-to-treat analysis; and sponsor involvement in authorship [72]. The Covidence software was used to manage data extraction and risk-of-bias assessments [73].

#### Data Synthesis and Analysis

The primary outcomes included pain, functioning, and quality of life, as assessed using recognized and validated measures [74]. Cost, emotional functioning, overall progress, range of motion, muscle strength, knowledge, self-management, and process-related outcomes were all secondary outcomes of interest.

A 3-step method was used to identify distinct subgroups of digital health interventions according to the domains of the RMIC. First, the appropriate number of clusters was determined through a hierarchical agglomerative clustering analysis using the Euclidean distance and average silhouette methods, which measures the quality of a cluster. We tested for outliers by using the cluster membership of the distance method, which indicates how well an observation fits into the cluster that it has been assigned to [75]. No outliers were identified based on the results of this analysis. Second, a nonhierarchical cluster analysis based on the k-means algorithm was performed to validate the results of the hierarchical procedure by using the initial cluster centroid number from hierarchical clustering as a starting point [76,77]. This method establishes the presence of clusters by determining the average of all the data points in a cluster.

The grouping of the clusters was evaluated by performing a principal component analysis (PCA), which required data normalization, and the eigenvalues were calculated and analyzed in a biplot graph [78,79]. Assumptions of the PCA were tested following the procedure described by Kassambara [80] (ie, linearity of the data, level of measurement, and outliers). Finally, the clusters were visually evaluated using cluster plots and PCA. To provide an interpretation of the cluster, the cluster means of the digital health interventions were applied.

We used DerSimonian and Laird random-effects models to summarize the treatment effects and expressed the results as standardized mean differences (SMDs) for continuous outcomes using different scales together with 95% CIs. The SMD calculations were based on the effect differences between the baseline and last follow-up assessment [81]. In the systematic review, we included relevant studies; for the meta- and subgroup analyses, at least three independent studies were required to justify the meta-analysis [82].

Heterogeneity in treatment effects between studies was assessed using the restricted maximum likelihood method ($I^2$) statistics, with $I^2$ values of 25%, 50%, and 75% corresponding to low, moderate, and high levels of heterogeneity, respectively [83]. Potential sources of statistical heterogeneity were explored using a priori subgroup analysis to determine whether the intervention duration (1-12 months or >12 months) or setting (clinic or home-based) affected heterogeneity. Evidence of small study effects was assessed through visual examination of funnel plots [84]. We conducted a sensitivity analysis of primary outcomes by excluding studies according to the following criteria: (1) high risk of bias, (2) long follow-up (≥12 months), and (3) large sample size (>200 participants). We used a minimum of 10 independent studies [81].

Descriptive statistics were used to summarize the data, where mean and SD were reported for continuous data and frequencies and percentages for categorical data. The distribution of all continuous variables was checked. The statistical significance for subgroup and sensitivity analysis was calculated using the test for subgroup differences provided in the R Studio (version 2021.09.01) package *meta*. All analyses were performed using the statistical software R Studio (Build 372), and libraries *dmetar*, *esc*, *tidyverse*, *meta*, *grid*, *robvis*, *pvclust*, and *factorextra* were used [85].

### Quality of Evidence

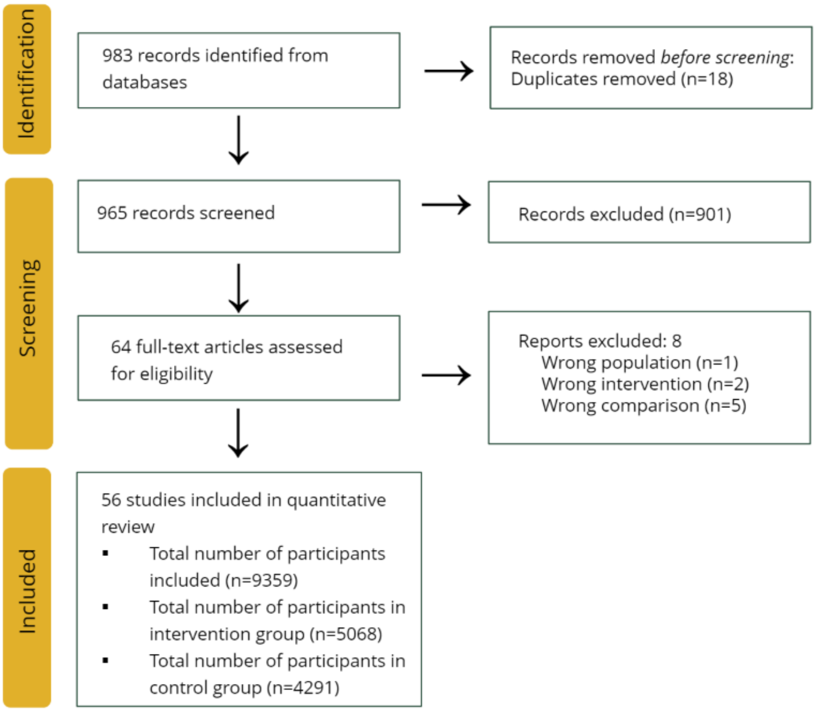
The quality of evidence was rated for each pooled analysis by using the Grades of Recommendation, Assessment, Development, and Evaluation approach [86]. The quality of evidence was not downgraded for performance or detection bias as perfect blinding is considered problematic for complex digital health interventions [82]. For each comparison, 2 researchers (LT and MAA) independently rated the quality of evidence for each outcome as "high," "moderate," "low," or "very low." Discrepancies were resolved through iteration and discussion.

## Results

### Search Results and Study Characteristics

A total of 983 publications of potential interest were identified. Of the 983 publications, after removing 18 (1.83%) duplicates, 965 (98.17%) publications were selected for title and abstract screening. Subsequently, of the 965 publications, 64 (6.63%) were selected for full-text screening, and 56 (5.8%) RCTs were considered eligible for inclusion, assessing 9359 participants. Approximately 6% (4/64) of studies reported incomplete outcomes; therefore, they were excluded from the effect analysis (Figure 2).

**Figure 2.** Flowchart of search strategy and study selection process.



### Intervention Characteristics

The characteristics of the interventions in the included studies are summarized in Tables S4 and S5 in Multimedia Appendix 1. All studies performed interventions in the home base of the participants, and in some cases, the setting of the study was a combination of the home base of the participants and the primary care clinic (26/56, 46%). In addition, of the 56 studies, we included 30 (54%) studies targeting musculoskeletal pain conditions, 10 (18%) studies targeting chronic pain conditions, 9 (16%) studies targeting postsurgery rehabilitation participants, and 7 (13%) studies focusing on patients with arthritis. Most of the interventions focused on patient conditions, such as target client communication (42/56, 75% studies; 6806/9359, 72.72% participants) and personal health tracking (38/56, 68% studies; 5881/9359, 62.84% participants). Digital health interventions at the professional level included telemedicine (55/56, 98% studies; 9331/9359, 99.7% participants), client information and registration (47/56, 84% studies; 8041/9359, 85.92% participants), health care provider decision support (23/56, 41% studies; 4520/9359, 48.3% participants), health care training (23/56, 41% studies; 4569/9359, 48.82% participants), health care provider communication (12/56, 21% studies; 2901/9359, 31% participants), and referral coordination (4/56, 7% studies; 527/9359, 5.63% participants). None of the studies incorporated health care providers in a scheduled activity planning intervention. Only 9% (5/56) of studies were targeted at the organizational level. All of these studies included health financing interventions (2363/9359, 25.25% participants). Furthermore, some studies were targeted at the system level and included data collection, management, and use interventions (23/56, 41% studies; 4648/9359, 49.66% participants). The duration of the interventions ranged from 2 weeks to 12 months (median 12 weeks). Two distinctive subgroups of digital health interventions were identified in the 56 articles.

The first cluster (32/56, 57% studies; 4565/9359, 48.78% participants) included interventions mainly in professional and client domains, mostly performed (23/32, 72%) in the home base of the participant. The second cluster (24/56, 43% studies; 4794/9359, 51.22% participants) comprised interventions in the organizational, professional, and client domains, mostly performed (17/24, 71%) in the home base and clinic settings. Four statistically significant differences across the subgroups for digital health interventions were identified through the cluster differences analysis: targeted health care provider decision assistance; referral coordination; health finance; and data collection, management, and use. The 2 clusters were named based on the characteristics of their digital health interventions (Table 1).

**Table 1.** Clusters of digital health interventions (N=56).

| Rainbow model intervention characterization | Total studies, n (%) | Cluster 1: patient-provider–level digital health interventions (n=32), n (%) | Cluster 2: patient-provider-organizational–level digital health interventions (n=24), n (%) | Cluster differences (*P* value) |
|---|---|---|---|---|
| **Organizational domain** | | | | |
| Health financing | 5 (9) | 0 (0) | 5 (21) | .006[a] |
| Data collection, management, and use | 23 (41) | 3 (9) | 20 (83) | <.001[b] |
| **Professional domain** | | | | |
| Client identification and registration | 47 (84) | 25 (78) | 22 (92) | .18 |
| Health care provider decision support | 23 (41) | 7 (22) | 16 (67) | <.001[c] |
| Telemedicine | 55 (98) | 32 (100) | 23 (96) | .25 |
| Health care provider communication | 12 (21) | 8 (25) | 4 (17) | .46 |
| Referral coordination | 4 (7) | 0 (0) | 4 (17) | .02[b] |
| Health care provider training | 23 (41) | 13 (41) | 10 (42) | .94 |
| **Client domain** | | | | |
| Targeted client communication | 42 (75) | 26 (81) | 16 (67) | .22 |
| Personal health tracking | 38 (68) | 21 (66) | 17 (71) | .69 |

[a] Significant at level .01.

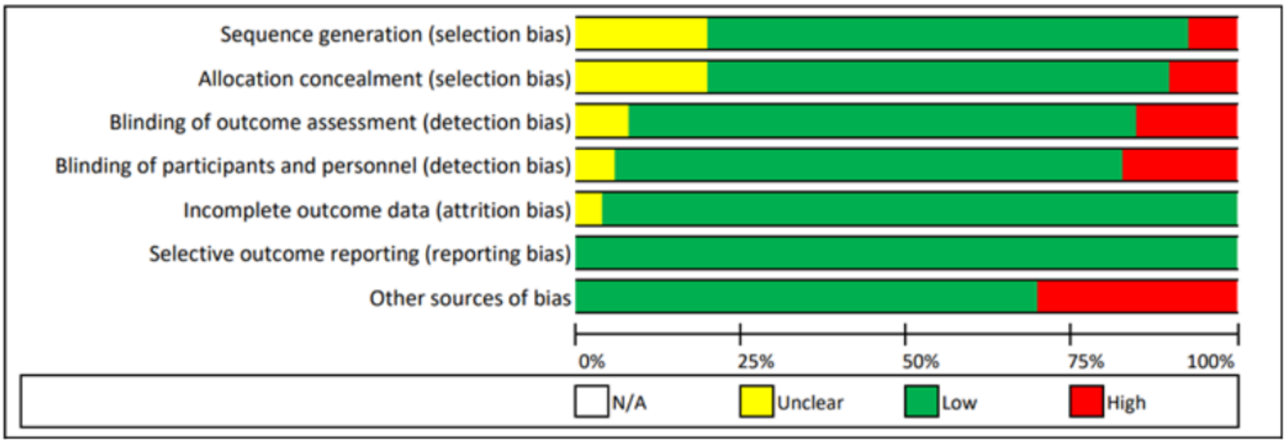
[b] Significant at level .001.

[c] Significant at level .05.

## Quality of Included Studies

The risk of bias in the included studies is summarized in Figure 3. Overall, there was a low risk of bias for 80.6% (316/392) of the items, an unclear risk for 8.7% (34/392), and a high risk for 10.7% (42/392).

**Figure 3.** Summary of the risks of bias in included studies. For each quality item, low risk means that sufficient data were reported in the study to allow the assessment of quality, and the study fulfilled the criteria for the quality item; high risk means that sufficient data were reported in the study to assess quality, but the study did not fulfill the criteria for the quality item; and unclear risk means that incomplete data for the quality item were reported. N/A: not applicable.



## Effect of Digital Health Interventions

### *Pain*

Of the 56 studies, 37 (66%; 5323/9359, 56.88% participants) reported the treatment effects on pain. Digital health interventions had a small effect on pain compared with standard care management (SMD 0.19, 95% CI 0.06-0.31; Figure 4). However, there was evidence of high heterogeneity between studies ($I^2$=81%). There was evidence of different effects on pain based on different types of digital health interventions (patient-provider: SMD 0.07, 95% CI −0.04 to 0.19; patient-provider-organization: SMD 0.34, 95% CI 0.08-0.60;

*P* value for subgroup difference=.05). The quality of the evidence for pain was rated as moderate (Table 2).

**Figure 4.** Effect of digital health on pain. SMD: standardized mean difference.

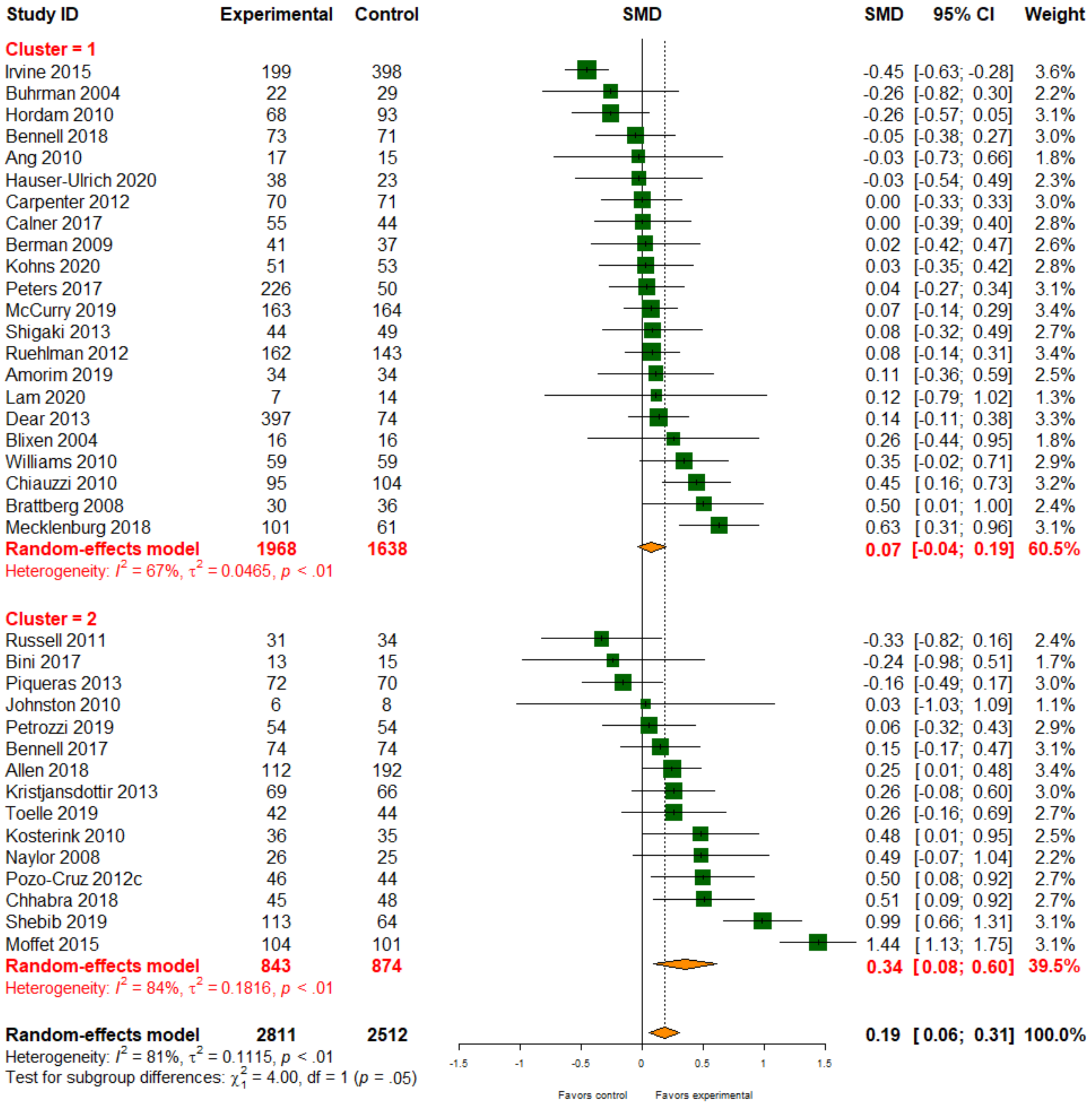


**Table 2.** Summary of findings and assessment of the quality of evidence for outcomes (N=56).

| Outcomes | Studies, n (%) | Certainty assessment | | | | | | Effect | | Certainty |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Study design | Risk of bias | Inconsistency | Indirectness | Imprecision | Other considerations | Individuals (n=9359), n (%) | SMD[a] rate (95% CI) | |
| Pain (follow-up: mean 25 weeks) | 37 (66.1) | Randomized trials | Serious[b] | Not serious | Not serious | Not serious | None | 5323 (56.9) | 0.19 (0.06 to 0.31) | Moderate |
| Disability and function (follow-up: mean 27 weeks) | 30 (53.6) | Randomized trials | Serious[b] | Not serious | Not serious | Not serious | None | 4849 (51.8) | 0.14 (0.03 to 0.25) | Moderate |
| Quality of life (follow-up: mean 25 weeks) | 24 (42.9) | Randomized trials | Not serious | Not serious | Not serious | Not serious | None | 3995 (42.5) | 0.22 (0.07 to 0.36) | High |
| Emotional functioning (follow-up: mean 29 weeks) | 24 (42.9) | Randomized trials | Serious[b] | Serious[c] | Not serious | Not serious | None | 3814 (40.8) | 0.24 (0.12 to 0.35) | Low |
| Self-management (follow-up: mean 26 weeks) | 21 (37.5) | Randomized trials | Serious[b] | Not serious | Not serious | Not serious | None | 2857 (30.5) | 0.14 (0.05 to 0.24) | Moderate |
| Global improvement (follow-up: mean 42 weeks) | 4 (7.1) | Randomized trials | Serious[b] | Not serious | Not serious | Serious[d] | None | 795 (5.5) | 0.25 (−0.44 to 0.93) | Low |

[a] SMD: standardized mean difference.

[b] Most of the studies had a high frequency of other bias.

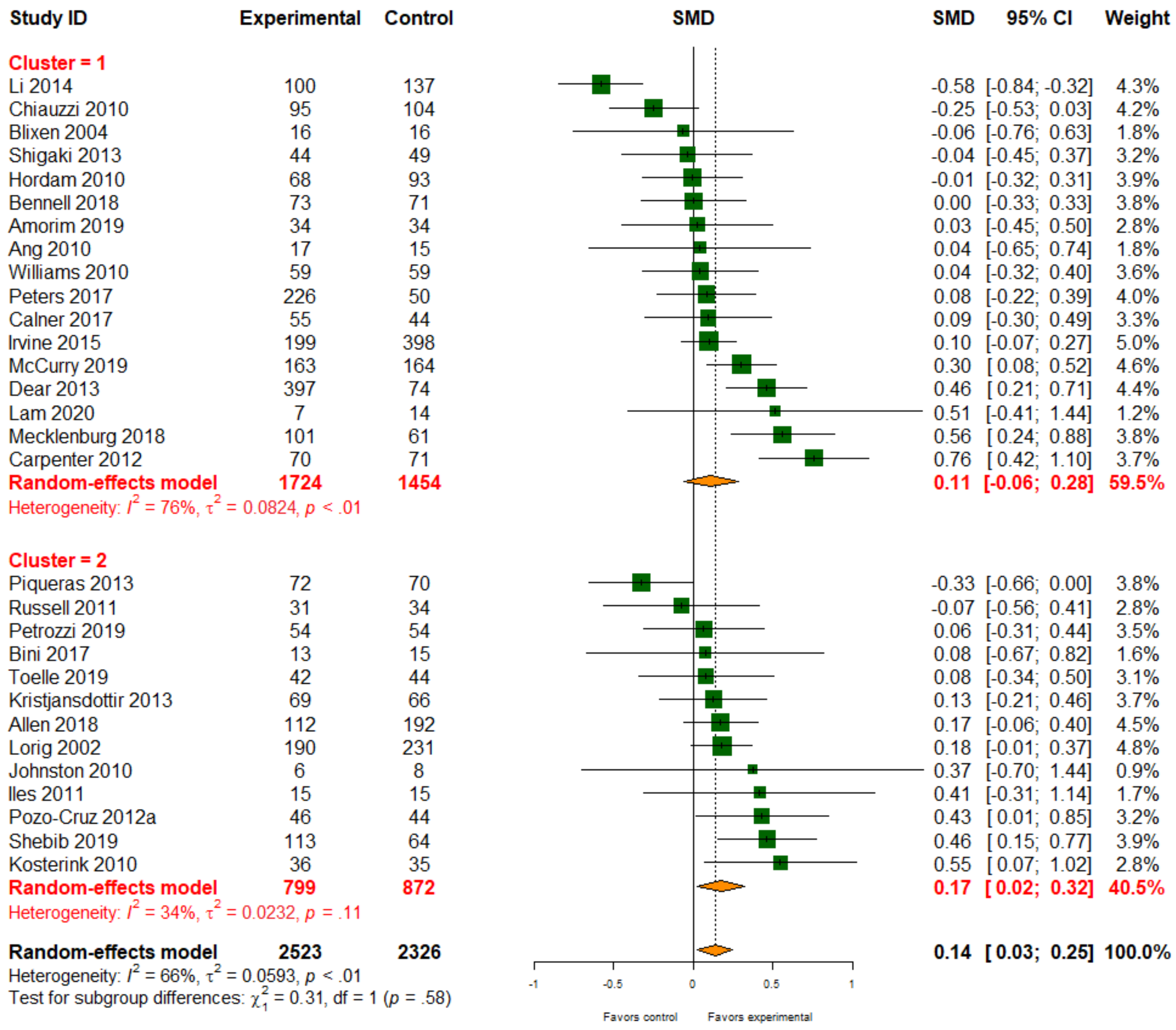
[c] Large heterogeneity between studies ($I^2$>50%).

[d] 95% CI includes the possible benefits from both control and digital health interventions.

### *Disability and Function*

Of the 56 studies, data on disability and function were reported in 30 (54%) studies (4849/9359, 51.8% participants). Digital health interventions slightly improved the functioning of people with musculoskeletal conditions (SMD 0.14, 95% CI 0.03-0.25); however, there was considerable heterogeneity among studies ($I^2$=66%; Figure 5). There was little evidence that different types of digital health interventions affected treatment effectiveness (patient-provider: SMD 0.11, 95% CI −0.06 to 0.28; patient-provider-organization: SMD 0.17, 95% CI 0.02-0.32; *P* value for subgroup difference=.58). The quality of the evidence for disability and functional outcomes was moderate.

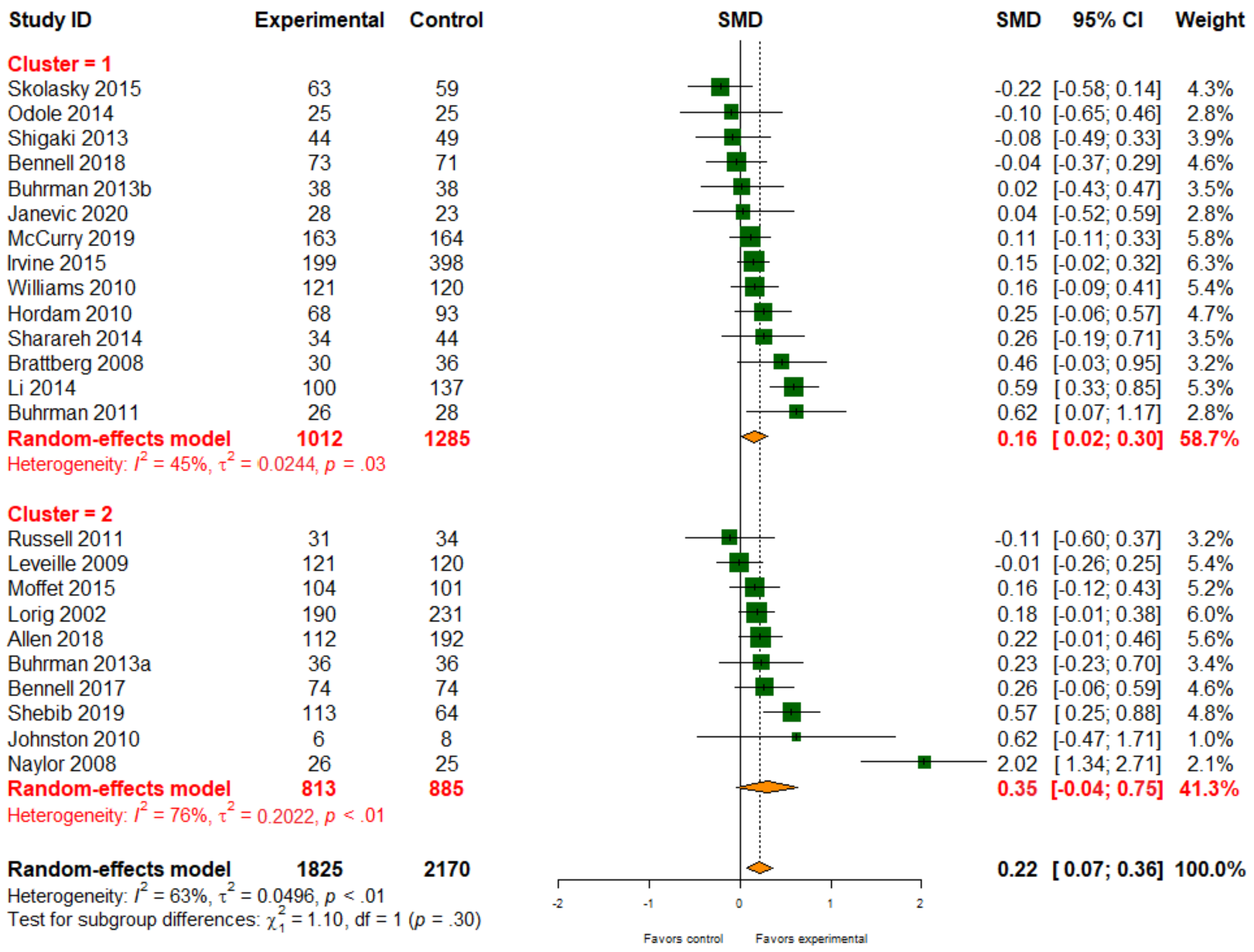
**Figure 5.** Effect of digital health on disability and function. SMD: standardized mean difference.



## Quality of Life

Digital health interventions had a slightly positive effect on health-related quality of life (24/56, 43% studies; 3995/9359, 42.69% participants; SMD 0.22, 95% CI 0.07-1.36). There was evidence of high-level heterogeneity between studies ($I^2$=63%; Figure 6). There was little evidence that different types of digital health interventions had differing effects on quality of life (patient-provider: SMD 0.16, 95% CI 0.02-0.30; patient-provider-organization: SMD 0.35, 95% CI −0.04 to 0.75; *P* value for subgroup difference=.30). The quality of evidence for the quality of life was graded as high.

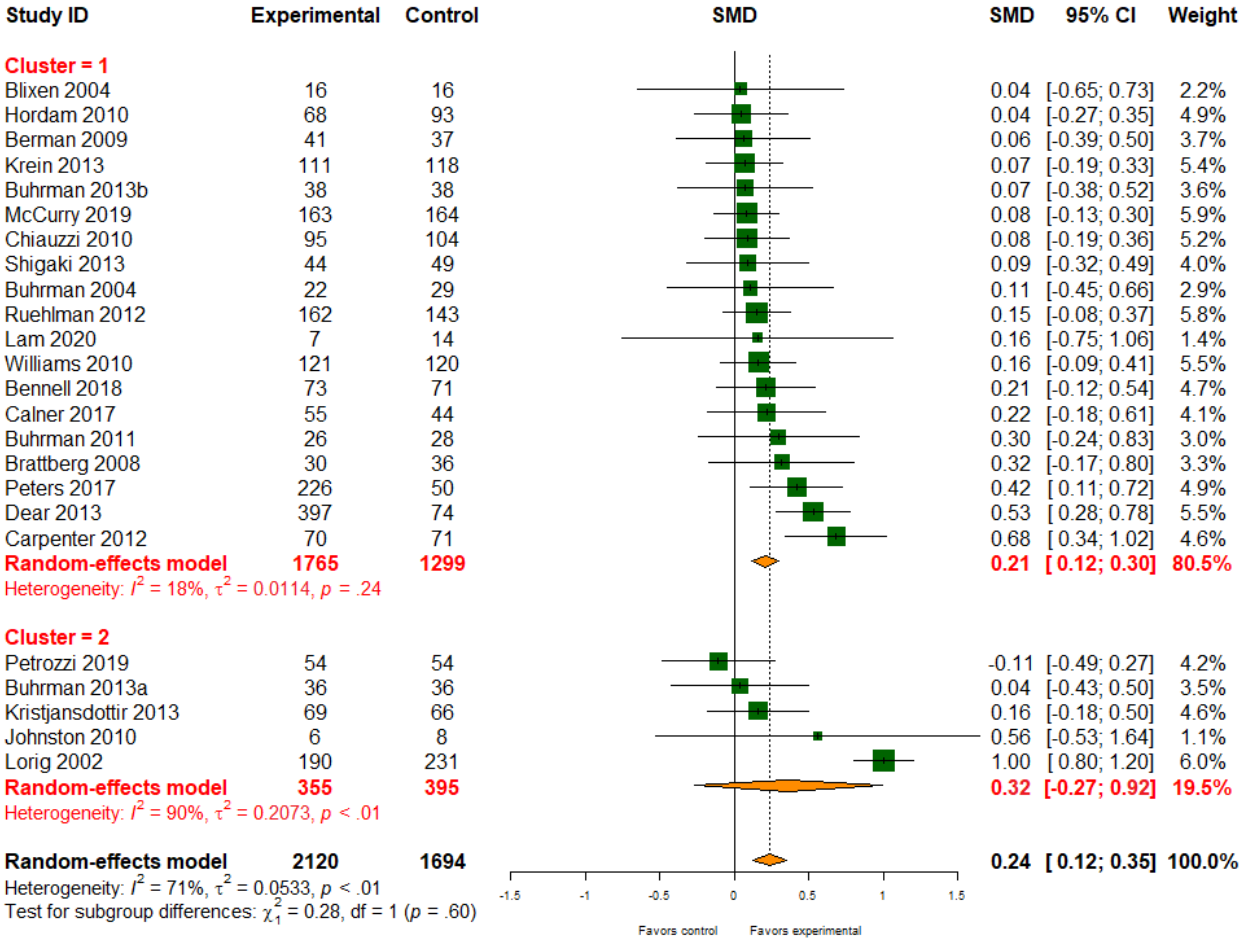
**Figure 6.** Effect of digital health on quality of life. SMD: standardized mean difference.



## *Emotional Functioning*

Of the 56 studies, 24 (43%; 3814/9359, 40.75% participants) reported data on emotional functioning. Digital health interventions had a positive effect on emotional functioning compared with usual care (SMD 0.24, 95% CI 0.12-0.35); however, there was evidence of heterogeneity between studies ($I^2$=71%; Figure 7). There was little evidence of different treatment effects for different types of interventions (patient-provider: SMD 0.21, 95% CI 0.12-0.30; patient-provider-organization: SMD 0.32, 95% CI −0.27 to 0.92; *P* value for subgroup difference=.60). The quality of evidence for emotional functioning was low.

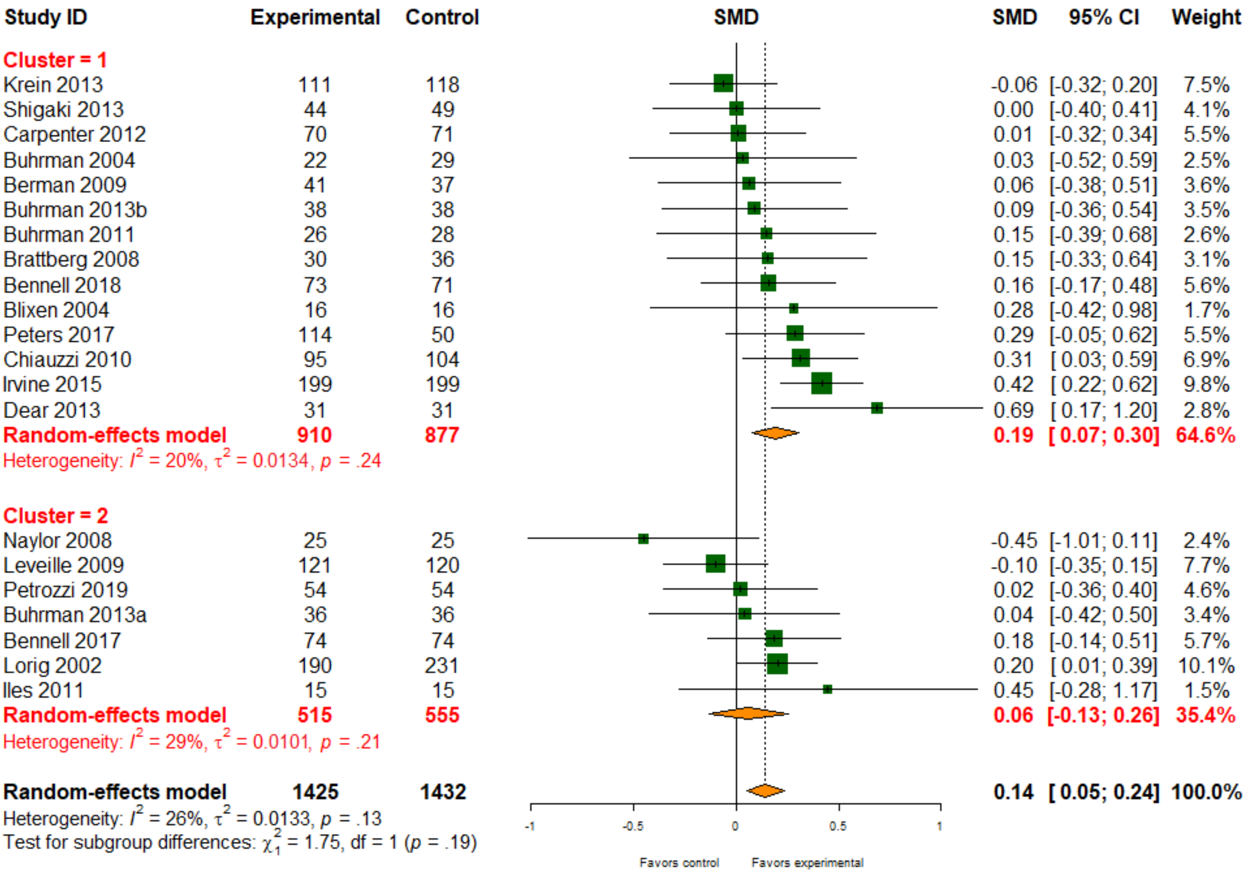
**Figure 7.** Effect of digital health on emotional functioning. SMD: standardized mean difference.



### Self-management

Of the 56 studies, 21 (38%) reported treatment effects on self-management behavior (2857/9359, 30.5% participants). Evidence suggests that digital health interventions have a small positive effect on self-management behaviors compared with usual care (SMD 0.14, 95% CI 0.05-0.24; Figure 8) with moderate quality of evidence. There was little evidence that different types of interventions affected treatment effectiveness (patient-provider: SMD 0.19, 95% CI 0.07-0.30; patient-provider-organization: SMD 0.14, 95% CI −0.13 to 0.26; *P* value for subgroup difference=.19).

**Figure 8.** Effect of digital health on self-management. SMD: standardized mean difference.



## Qualitative Synthesis

The qualitative analysis showed that digital health interventions have little or no effect on global improvement compared with standard care management (4/56, 7% studies, 795/9359, 8.49% participants; SMD 0.25, 95% CI −0.44 to 1.93). There was evidence of heterogeneity between studies ($I^2$=87%), with a low quality of evidence. In addition, data on the range of motion were provided from 4% (2/56) of investigations involving 2.24% (210/9359) of participants; however, the treatment effects were highly ambiguous (Table S6 in Multimedia Appendix 1 [16-71]). Furthermore, 4% (2/56) of studies reported no effect of digital health on muscle strength (Table S6 in Multimedia Appendix 1 [16-71]). Of the 56 studies, the effects of digital health interventions on knowledge were reported in 2 (4%) studies (774/9359, 8.27% participants), and 1 (2%) study reported a significant effect (Table S6 in Multimedia Appendix 1 [16-71]). One of the studies reported an effect on satisfaction scores among participants, and another reported recovery expectation rates during the intervention (Table S6 in Multimedia Appendix 1 [16-71]). A cost analysis of digital health interventions for individuals with musculoskeletal pain conditions was presented in 4% (2/56) of studies (349/9359, 3.73% participants). In both investigations, digital health interventions were cost-effective and efficient (Table S6 in Multimedia Appendix 1 [16-71]).

### Publication Bias, Subgroup, and Sensitivity Analyses

There was little evidence of funnel plot asymmetry in treatment effects for pain, disability and function, quality of life, and emotional functioning (Figures S1-S5 in Multimedia Appendix 1 [16-71]). In addition, there was little evidence that digital health interventions had different effects on pain, disability and function, quality of life, emotional functioning, and self-management based on the duration of intervention (pain *P*=.66; disability and function *P*=.94; quality of life *P*=.45; emotional functioning *P*=.42; and self-management *P*=.66) or study setting (pain *P*=.80; disability and function *P*=.05; quality of life *P*=.63; emotional functioning *P*=.06; and self-management *P*=.06). The sensitivity analysis showed that restricting analyses to studies with lower risks of bias (pain *P*=.15; disability and function *P*=.58; quality of life *P*=.26; and self-management *P*=.39), follow-up <12 months (pain *P*=.22; disability and function *P*=.66; quality of life *P*=.31; emotional functioning *P*=.85; and self-management *P*=.48), or a small sample size (pain *P*=.88; disability and function *P*=.74; quality of life *P*=.62; emotional functioning, *P*=.19; and self-management *P*=.85) provided no different treatment effects for pain, disability and function, quality of life, and self-management (Table S7 in Multimedia Appendix 1 [16-71]). However, the risk of bias resulted in different results for emotional functioning (*P*=.01).

## Discussion

### Principal Findings

To the best of our knowledge, this meta-analytic review is the first to systematically assess the effectiveness of digital health interventions among people with musculoskeletal pain conditions. Pain, functioning, quality of life, emotional functioning, and self-management were all found to have small positive effects on a diverse set of digital health interventions. There was evidence that multicomponent interventions targeted at the client, provider, and organization levels had greater effects on pain than interventions targeted only at the client and provider levels. There was little evidence that different types of digital health interventions had different effects on other outcomes. The lack of high-quality evidence on global improvement, range of motion, muscle strength, and knowledge reinforces the need for further research on digital health for musculoskeletal pain conditions.

## Comparison With Existing Evidence

Previous reviews have also reported evidence on the effects of digital health interventions for reducing pain in musculoskeletal conditions [4,5,87-89]. However, most of these studies focused solely on chronic pain [87,88] or generic musculoskeletal conditions [5,7]. Further research is needed to corroborate our findings linking compound digital health treatments at the patient, provider, and organizational levels to reduced pain symptoms.

Our findings highlighting that digital health interventions improve function are consistent with earlier reviews of studies involving patients with generic musculoskeletal conditions [5,7]. Reviews focusing on chronic and nonspecific low back pain populations have reported limited evidence on the effects of digital health interventions on improving function [4,88]. The complexity of (chronic) pain management and the small number of RCTs included in earlier evaluations could explain the disparity in results.

This review indicates that digital health interventions have little effect on health-related quality of life. Previous systematic reviews have been inconsistent in this regard. For instance, 2 reviews suggested nonsignificant quality of life effects on musculoskeletal and chronic pain conditions [7,87], whereas 1 review reported a significant improvement in quality of life among people with nonspecific low back pain [4]. The variability of results may be explained by the differences in target populations, quality of the study design, and number of RCTs included in previous studies.

Similarly, this study has shown favorable outcomes for the emotional functioning of digital health interventions for people with musculoskeletal pain [87,88]. However, the sensitivity analysis provides evidence of the risk of bias confounding the effects, which requires further investigation. In line with other studies, this review found that digital health interventions may increase self-management behavior [88].

In all the reviewed studies, there was only a minimal reference to the cost-effectiveness of digital health interventions for musculoskeletal pain conditions. We could only include 2 studies reporting a significant cost reduction of digital health interventions compared with usual care [16,17]. Future trials should further explore whether digital health interventions can improve health outcomes related to musculoskeletal pain at lower costs than usual care. Data reporting for global improvement, range of motion, muscle strength, knowledge, and the delivery process of digital health were notably underreported, as has been observed in other reviews [4,5,87-89].

## Strengths and Limitations

This is the first review to synthesize the types of digital health interventions reported in the literature and quantify their effectiveness and confidence in treatment effects across a broad range of outcome measures. The strength of this review is that it was theoretically grounded in the WHO taxonomy [11] and the RMIC [13] to classify ambiguous digital health interventions reported in the literature. However, some limitations of this study must be acknowledged. First, it must be noted that confounding factors carry an inherent risk of bias, as evidenced by the large statistical heterogeneity across the pooled results for pain, function, quality of life, and self-management. In addition, the effects found in this study could have been influenced by differences in measurement scales and not by real differences in variability among study populations [90,91]. This should be further investigated in future studies. Moreover, the content of digital health interventions, diagnostic groups, and control conditions varied considerably, potentially biasing the results. Therefore, generalizing the overall findings to the management of musculoskeletal pain conditions should be treated with caution. Second, although we used a broad search technique, this evaluation could have been hindered by language bias, as we only included English-language literature. This means that our search may not reflect all available digital health interventions for musculoskeletal pain conditions. Third, we did not find any evidence of publication bias. It should be noted that the Egger test could potentially be misleading when used with continuous outcome measures [92]. Finally, although we abstracted and summarized the essential components of the interventions, there was minimal information on the type and intensity of digital health interventions offered.

## Relevance for Clinical Practice and Research

A major finding was that digital health interventions targeted at the clinical, provider, and organizational levels were effective in reducing pain for musculoskeletal conditions. To date, most studies have focused on isolated digital interventions targeted at the patient-provider level, such as telemedicine or targeted client communication. Future research should focus on improving the longitudinal design and on different types of interventions, drawing on the recent WHO taxonomy and the RMIC. Our findings should encourage interest in implementing real-world evaluation designs of digital health models to improve health care delivery as digital health interventions become more prevalent. Moreover, none of the studies included in this review covered the full breadth of the triple aim of assessing health, quality of care, and cost outcomes in conjunction. This emphasizes the importance of creating a core triple-aim result set for digital health interventions, which includes a defined set of outcomes that measure user experience, intervention quality, and costs.

## Conclusions

This review provides moderate-quality evidence that digital health interventions are effective in reducing pain and improving functioning and self-management of musculoskeletal pain conditions. Low-quality evidence indicates that digital health can improve the quality of life and global treatment. Although evaluations of the effects of digital health on costs, knowledge, global improvement, range of motion, muscle strength, and implementation fidelity are limited, these findings point to the need for more primary research into the particular combination of digital interventions that health care providers could use effectively.

## Acknowledgments

This work was supported by a research grant from the Nordic Netcare Group (Copenhagen, Denmark). TJ, JK, and CWB are employees of the Nordic Netcare Group.

## Authors' Contributions

PPV, LT, and RYA conceptualized and designed the study; PPV, LT, and MAA acquired, analyzed, and interpreted the data; PPV drafted the manuscript; PPV, LT, MAA, TJ, JK, CWB, and RYA critically revised the manuscript for important intellectual content; LT and PPV conducted statistical analyses; PPV and LT supervised the study. All authors read and approved the final manuscript. PPV and LT had full access to all data in the study and take responsibility for the integrity of the data and accuracy of the data analysis.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures of search strategy and study results.
[DOCX File , 425 KB - jmir_v24i9e37869_app1.docx ]

## References

1. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: a systematic analysis of the World Health Organization Burden of Diseases Database. Ann Rheum Dis 2019 Jun;78(6):844-848. [doi: 10.1136/annrheumdis-2019-215142] [Medline: 30987966]
2. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 09;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
3. Briggs AM, Towler SC, Speerin R, March LM. Models of care for musculoskeletal health in Australia: now more than ever to drive evidence into health policy and practice. Aust Health Rev 2014 Sep;38(4):401-405. [doi: 10.1071/AH14032] [Medline: 25086678]
4. Dario AB, Moreti Cabral A, Almeida L, Ferreira ML, Refshauge K, Simic M, et al. Effectiveness of telehealth-based interventions in the management of non-specific low back pain: a systematic review with meta-analysis. Spine J 2017 Sep;17(9):1342-1351. [doi: 10.1016/j.spinee.2017.04.008] [Medline: 28412562]
5. 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 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
6. Jiang S, Xiang J, Gao X, Guo K, Liu B. The comparison of telerehabilitation and face-to-face rehabilitation after total knee arthroplasty: a systematic review and meta-analysis. J Telemed Telecare 2018 May;24(4):257-262. [doi: 10.1177/1357633X16686748] [Medline: 28027679]
7. Cottrell MA, Galea OA, O'Leary SP, Hill AJ, Russell TG. Real-time telerehabilitation for the treatment of musculoskeletal conditions is effective and comparable to standard practice: a systematic review and meta-analysis. Clin Rehabil 2017 May;31(5):625-638. [doi: 10.1177/0269215516645148] [Medline: 27141087]
8. Slattery BW, Haugh S, Francis K, O'Connor L, Barrett K, Dwyer CP, et al. Protocol for a systematic review with network meta-analysis of the modalities used to deliver eHealth interventions for chronic pain. Syst Rev 2017 Mar 03;6(1):45 [FREE Full text] [doi: 10.1186/s13643-017-0414-x] [Medline: 28253909]
9. Heapy AA, Higgins DM, Cervone D, Wandner L, Fenton BT, Kerns RD. A systematic review of technology-assisted self-management interventions for chronic pain: looking across treatment modalities. Clin J Pain 2015 Jun;31(6):470-492. [doi: 10.1097/AJP.0000000000000185] [Medline: 25411862]
10. Kelly M, Fullen B, Martin D, McMahon S, McVeigh JG. eHealth interventions to support self-management in people with musculoskeletal disorders: a scoping review protocol. JBI Evid Synth 2021 Mar;19(3):709-720. [doi: 10.11124/JBIES-20-00147] [Medline: 33725715]

11. Recommendations on digital interventions for health system strengthening. World Health Organization. Geneva, Switzerland: World Health Organization; 2019 Jun 6. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2021-07-06]
12. Asthana S, Jones R, Sheaff R. Why does the NHS struggle to adopt eHealth innovations? A review of macro, meso and micro factors. BMC Health Serv Res 2019 Dec 21;19(1):984 [FREE Full text] [doi: 10.1186/s12913-019-4790-x] [Medline: 31864370]
13. Valentijn PP, Schepman SM, Opheij W, Bruijnzeels MA. Understanding integrated care: a comprehensive conceptual framework based on the integrative functions of primary care. Int J Integr Care 2013 Mar 22;13:e010 [FREE Full text] [doi: 10.5334/ijic.886] [Medline: 23687482]
14. Rainbow Model for Integrated Care. Essenburgh Group. 2021. URL: https://www.essenburgh.com/en/rainbow-model-of-integrated-care [accessed 2021-07-06]
15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
16. Paganini S, Lin J, Kählke F, Buntrock C, Leiding D, Ebert DD, et al. A guided and unguided Internet- and mobile-based intervention for chronic pain: health economic evaluation alongside a randomised controlled trial. BMJ Open 2019 Apr 09;9(4):e023390 [FREE Full text] [doi: 10.1136/bmjopen-2018-023390] [Medline: 30967405]
17. Fatoye F, Gebrye T, Fatoye C, Mbada CE, Olaoye MI, Odole AC, et al. The clinical and cost-effectiveness of telerehabilitation for people with nonspecific chronic low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 24;8(6):e15375 [FREE Full text] [doi: 10.2196/15375] [Medline: 32357128]
18. Brattberg G. Self-administered EFT (Emotional Freedom Techniques) in individuals with fibromyalgia: a randomized trial. Integr Med 2008;7(4):30-35.
19. Allen KD, Arbeeva L, Callahan LF, Golightly YM, Goode AP, Heiderscheit BC, et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Osteoarthritis Cartilage 2018 Mar;26(3):383-396 [FREE Full text] [doi: 10.1016/j.joca.2017.12.008] [Medline: 29307722]
20. Lorig KR, Laurent DD, Deyo RA, Marnell ME, Minor MA, Ritter PL. Can a Back Pain E-mail Discussion Group improve health status and lower health care costs?: a randomized study. Arch Intern Med 2002 Apr 08;162(7):792-796. [doi: 10.1001/archinte.162.7.792] [Medline: 11926853]
21. Buhrman M, Fredriksson A, Edström G, Shafiei D, Tärnqvist C, Ljótsson B, et al. Guided Internet-delivered cognitive behavioural therapy for chronic pain patients who have residual symptoms after rehabilitation treatment: randomized controlled trial. Eur J Pain 2013 May;17(5):753-765. [doi: 10.1002/j.1532-2149.2012.00244.x] [Medline: 23139021]
22. Hørdam B, Sabroe S, Pedersen PU, Mejdahl S, Søballe K. Nursing intervention by telephone interviews of patients aged over 65 years after total hip replacement improves health status: a randomised clinical trial. Scand J Caring Sci 2010 Mar;24(1):94-100. [doi: 10.1111/j.1471-6712.2009.00691.x] [Medline: 19422632]
23. Buhrman M, Skoglund A, Husell J, Bergström K, Gordh T, Hursti T, et al. Guided internet-delivered acceptance and commitment therapy for chronic pain patients: a randomized controlled trial. Behav Res Ther 2013 Jun;51(6):307-315. [doi: 10.1016/j.brat.2013.02.010] [Medline: 23548250]
24. Bennell KL, Nelligan RK, Rini C, Keefe FJ, Kasza J, French S, et al. Effects of internet-based pain coping skills training before home exercise for individuals with hip osteoarthritis (HOPE trial): a randomised controlled trial. Pain 2018 Sep;159(9):1833-1842. [doi: 10.1097/j.pain.0000000000001281] [Medline: 29794609]
25. Berman RL, Iris MA, Bode R, Drengenberg C. The effectiveness of an online mind-body intervention for older adults with chronic pain. J Pain 2009 Jan;10(1):68-79. [doi: 10.1016/j.jpain.2008.07.006] [Medline: 18774342]
26. Buhrman M, Fältenhag S, Ström L, Andersson G. Controlled trial of Internet-based treatment with telephone support for chronic back pain. Pain 2004 Oct;111(3):368-377. [doi: 10.1016/j.pain.2004.07.021] [Medline: 15363881]
27. Peters ML, Smeets E, Feijge M, van Breukelen G, Andersson G, Buhrman M, et al. Happy despite pain: a randomized controlled trial of an 8-week Internet-delivered positive psychology intervention for enhancing well-being in patients with chronic pain. Clin J Pain 2017 Nov;33(11):962-975 [FREE Full text] [doi: 10.1097/AJP.0000000000000494] [Medline: 28379873]
28. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
29. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
30. Calner T, Nordin C, Eriksson MK, Nyberg L, Gard G, Michaelson P. Effects of a self-guided, Web-based activity programme for patients with persistent musculoskeletal pain in primary healthcare: a randomized controlled trial. Eur J Pain 2017 Jul;21(6):1110-1120. [doi: 10.1002/ejp.1012] [Medline: 28464364]
31. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]

32. Johnston M, Foster M, Shennan J, Starkey NJ, Johnson A. The effectiveness of an acceptance and commitment therapy self-help intervention for chronic pain. Clin J Pain 2010 Jun;26(5):393-402. [doi: 10.1097/AJP.0b013e3181cf59ce] [Medline: 20473046]
33. Blixen CE, Bramstedt KA, Hammel JP, Tilley BC. A pilot study of health education via a nurse-run telephone self-management programme for elderly people with osteoarthritis. J Telemed Telecare 2004;10(1):44-49. [doi: 10.1258/135763304322764194] [Medline: 15006216]
34. Ang DC, Chakr R, Mazzuca S, France CR, Steiner J, Stump T. Cognitive-behavioral therapy attenuates nociceptive responding in patients with fibromyalgia: a pilot study. Arthritis Care Res (Hoboken) 2010 May;62(5):618-623 [FREE Full text] [doi: 10.1002/acr.20119] [Medline: 20191481]
35. Shigaki CL, Smarr KL, Siva C, Ge B, Musser D, Johnson R. RAHelp: an online intervention for individuals with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2013 Oct;65(10):1573-1581 [FREE Full text] [doi: 10.1002/acr.22042] [Medline: 23666599]
36. Petrozzi MJ, Leaver A, Ferreira PH, Rubinstein SM, Jones MK, Mackey MG. Addition of MoodGYM to physical treatments for chronic low back pain: a randomized controlled trial. Chiropr Man Therap 2019 Oct 25;27:54 [FREE Full text] [doi: 10.1186/s12998-019-0277-4] [Medline: 31673330]
37. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019 Jan 7;2:1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
38. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019 May 3;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
39. Odole AC, Ojo OD. Is telephysiotherapy an option for improved quality of life in patients with osteoarthritis of the knee? Int J Telemed Appl 2014;2014:903816 [FREE Full text] [doi: 10.1155/2014/903816] [Medline: 24778645]
40. Krein SL, Kadri R, Hughes M, Kerr EA, Piette JD, Holleman R, et al. Pedometer-based Internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. J Med Internet Res 2013 Aug 19;15(8):e181 [FREE Full text] [doi: 10.2196/jmir.2605] [Medline: 23969029]
41. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
42. Buhrman M, Nilsson-Ihrfeldt E, Jannert M, Ström L, Andersson G. Guided internet-based cognitive behavioural treatment for chronic back pain reduces pain catastrophizing: a randomized controlled trial. J Rehabil Med 2011 May;43(6):500-505 [FREE Full text] [doi: 10.2340/16501977-0805] [Medline: 21533329]
43. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an Internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 04;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
44. Leveille SG, Huang A, Tsai SB, Allen M, Weingart SN, Iezzoni LI. Health coaching via an Internet portal for primary care patients with chronic conditions: a randomized controlled trial. Med Care 2009 Jan;47(1):41-47. [doi: 10.1097/MLR.0b013e3181844dd0] [Medline: 19106729]
45. Williams DA, Kuper D, Segar M, Mohan N, Sheth M, Clauw DJ. Internet-enhanced management of fibromyalgia: a randomized controlled trial. Pain 2010 Dec;151(3):694-702 [FREE Full text] [doi: 10.1016/j.pain.2010.08.034] [Medline: 20855168]
46. Allen KD, Oddone EZ, Coffman CJ, Datta SK, Juntilla KA, Lindquist JH, et al. Telephone-based self-management of osteoarthritis: a randomized trial. Ann Intern Med 2010 Nov 02;153(9):570-579 [FREE Full text] [doi: 10.7326/0003-4819-153-9-201011020-00006] [Medline: 21041576]
47. Ruehlman LS, Karoly P, Enders C. A randomized controlled evaluation of an online chronic pain self management program. Pain 2012 Feb;153(2):319-330 [FREE Full text] [doi: 10.1016/j.pain.2011.10.025] [Medline: 22133450]
48. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
49. Dear BF, Titov N, Perry KN, Johnston L, Wootton BM, Terides MD, et al. The Pain Course: a randomised controlled trial of a clinician-guided Internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. Pain 2013 Jun;154(6):942-950. [doi: 10.1016/j.pain.2013.03.005] [Medline: 23688830]
50. Naylor MR, Keefe FJ, Brigidi B, Naud S, Helzer JE. Therapeutic Interactive Voice Response for chronic pain reduction and relapse prevention. Pain 2008 Feb;134(3):335-345 [FREE Full text] [doi: 10.1016/j.pain.2007.11.001] [Medline: 18178011]
51. Del Pozo-Cruz B, Adsuar JC, Parraca J, Del Pozo-Cruz J, Moreno A, Gusi N. A Web-based intervention to improve and prevent low back pain among office workers: a randomized controlled trial. J Orthop Sports Phys Ther 2012 Oct;42(10):831-841. [doi: 10.2519/jospt.2012.3980] [Medline: 22951407]
52. Kristjánsdóttir Ó, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain.

part 2: 11-month follow-up results of a randomized trial. J Med Internet Res 2013 Mar 28;15(3):e72 [FREE Full text] [doi: 10.2196/jmir.2442] [Medline: 23538392]

53. Janevic MR, Shute V, Murphy SL, Piette JD. Acceptability and effects of commercially available activity trackers for chronic pain management among older African American adults. Pain Med 2020 Feb 01;21(2):e68-e78 [FREE Full text] [doi: 10.1093/pm/pnz215] [Medline: 31509196]
54. McCurry SM, Von Korff M, Morin CM, Cunningham A, Pike KC, Thakral M, et al. Telephone interventions for co-morbid insomnia and osteoarthritis pain: the OsteoArthritis and Therapy for Sleep (OATS) randomized trial design. Contemp Clin Trials 2019 Dec;87:105851 [FREE Full text] [doi: 10.1016/j.cct.2019.105851] [Medline: 31614214]
55. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
56. Kohns DJ, Urbanik CP, Geisser ME, Schubiner H, Lumley MA. The effects of a pain psychology and neuroscience self-evaluation internet intervention: a randomized controlled trial. Clin J Pain 2020 Sep;36(9):683-692. [doi: 10.1097/AJP.0000000000000857] [Medline: 32520816]
57. Lam J, Svensson P, Alstergren P. Internet-based multimodal pain program with telephone support for adults with chronic temporomandibular disorder pain: randomized controlled pilot trial. J Med Internet Res 2020 Oct 13;22(10):e22326 [FREE Full text] [doi: 10.2196/22326] [Medline: 33048053]
58. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating Mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. BMC Musculoskelet Disord 2019 Feb 11;20(1):71 [FREE Full text] [doi: 10.1186/s12891-019-2454-y] [Medline: 30744606]
59. Skolasky RL, Maggard AM, Li D, Riley 3rd LH, Wegener ST. Health behavior change counseling in surgery for degenerative lumbar spinal stenosis. Part I: improvement in rehabilitation engagement and functional outcomes. Arch Phys Med Rehabil 2015 Jul;96(7):1200-1207 [FREE Full text] [doi: 10.1016/j.apmr.2015.03.009] [Medline: 25827657]
60. Sharareh B, Schwarzkopf R. Effectiveness of telemedical applications in postoperative follow-up after total joint arthroplasty. J Arthroplasty 2014 May;29(5):918-22.e1. [doi: 10.1016/j.arth.2013.09.019] [Medline: 24342278]
61. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. In-home telerehabilitation compared with face-to-face rehabilitation after total knee arthroplasty: a noninferiority randomized controlled trial. J Bone Joint Surg Am 2015 Jul 15;97(14):1129-1141. [doi: 10.2106/JBJS.N.01066] [Medline: 26178888]
62. Piqueras M, Marco E, Coll M, Escalada F, Ballester A, Cinca C, et al. Effectiveness of an interactive virtual telerehabilitation system in patients after total knee arthoplasty: a randomized controlled trial. J Rehabil Med 2013 Apr;45(4):392-396 [FREE Full text] [doi: 10.2340/16501977-1119] [Medline: 23474735]
63. Bini SA, Mahajan J. Clinical outcomes of remote asynchronous telerehabilitation are equivalent to traditional therapy following total knee arthroplasty: a randomized control study. J Telemed Telecare 2017 Feb;23(2):239-247. [doi: 10.1177/1357633X16634518] [Medline: 26940798]
64. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: a randomized controlled trial. J Bone Joint Surg Am 2011 Jan 19;93(2):113-120. [doi: 10.2106/JBJS.I.01375] [Medline: 21248209]
65. Li LL, Gan YY, Zhang LN, Wang YB, Zhang F, Qi JM. The effect of post-discharge telephone intervention on rehabilitation following total hip replacement surgery. Int J Nurs Sci 2014 Jun;1(2):207-211. [doi: 10.1016/j.ijnss.2014.05.005]
66. Tousignant M, Moffet H, Boissy P, Corriveau H, Cabana F, Marquis F. A randomized controlled trial of home telerehabilitation for post-knee arthroplasty. J Telemed Telecare 2011;17(4):195-198. [doi: 10.1258/jtt.2010.100602] [Medline: 21398389]
67. Iles R, Taylor NF, Davidson M, O'Halloran P. Telephone coaching can increase activity levels for people with non-chronic low back pain: a randomised trial. J Physiother 2011;57(4):231-238 [FREE Full text] [doi: 10.1016/S1836-9553(11)70053-4] [Medline: 22093121]
68. Priebe JA, Haas KK, Moreno Sanchez LF, Schoefmann K, Utpadel-Fischler DA, Stockert P, et al. Digital treatment of back pain versus standard of care: the cluster-randomized controlled trial, Rise-uP. J Pain Res 2020 Jul 17;13:1823-1838 [FREE Full text] [doi: 10.2147/JPR.S260761] [Medline: 32765057]
69. Kosterink SM, Huis in 't Veld RM, Cagnie B, Hasenbring M, Vollenbroek-Hutten MM. The clinical effectiveness of a myofeedback-based teletreatment service in patients with non-specific neck and shoulder pain: a randomized controlled trial. J Telemed Telecare 2010;16(6):316-321. [doi: 10.1258/jtt.2010.006005] [Medline: 20798425]
70. del Pozo-Cruz B, Gusi N, del Pozo-Cruz J, Adsuar JC, Hernandez-Mocholí M, Parraca JA. Clinical effects of a nine-month Web-based intervention in subacute non-specific low back pain patients: a randomized controlled trial. Clin Rehabil 2013 Jan;27(1):28-39. [doi: 10.1177/0269215512444632] [Medline: 22653374]
71. del Pozo-Cruz B, Parraca J, del Pozo-Cruz J, Adsuar J, Hill J, Gusi N. An occupational, Internet-based intervention to prevent chronicity in subacute lower back pain: a randomised controlled trial. J Rehabil Med 2012 Jun;44(7):581-587 [FREE Full text] [doi: 10.2340/16501977-0988] [Medline: 22674240]

72. 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]
73. Covidence: Accelerate Your Systematic Review. Covidence. 2017. URL: https://www.covidence.org/ [accessed 2021-07-06]
74. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
75. Piot M. Chapter 11: Clustering, distance methods and ordination. In: Statistics in Climate Sciences. Bern, Switzerland: Departement Mathematik und Statistik, Universität Bern; 2014:11-1-1112.
76. Hair JF. Multivariate Data Analysis: A Global Perspective. 7th edition. Hoboken, NJ, USA: Prentice Hall; 2009.
77. Valentijn PP, Ruwaard D, Vrijhoef HJ, de Bont A, Arends RY, Bruijnzeels MA. Collaboration processes and perceived effectiveness of integrated care projects in primary care: a longitudinal mixed-methods study. BMC Health Serv Res 2015 Oct 09;15:463 [FREE Full text] [doi: 10.1186/s12913-015-1125-4] [Medline: 26450573]
78. Leznik M, Tofallis C. Estimating invariant principal components using diagonal regression. University of Hertfordshire. 2005. URL: http://uhra.herts.ac.uk/handle/2299/715 [accessed 2022-05-11]
79. Gabriel KR. The biplot graphic display of matrices with application to principal component analysis. Biometrika 1971 Dec;58(3):453-467. [doi: 10.2307/2334381]
80. Kassambara A. Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning. Scotts Valley, CA, USA: CreateSpace Independent Publishing Platform; 2017.
81. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ, USA: John Wiley & Sons; 2019.
82. Valentijn PP, Pereira FA, Ruospo M, Palmer SC, Hegbrant J, Sterner CW, et al. Person-centered integrated care for chronic kidney disease: a systematic review and meta-analysis of randomized controlled trials. Clin J Am Soc Nephrol 2018 Mar 07;13(3):375-386 [FREE Full text] [doi: 10.2215/CJN.09960917] [Medline: 29438975]
83. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
84. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
85. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2013. URL: http://r.meteo.uni.wroc.pl/web/packages/dplR/vignettes/intro-dplR.pdf [accessed 2021-07-06]
86. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
87. Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, et al. An evaluation of the effectiveness of the modalities used to deliver electronic health interventions for chronic pain: systematic review with network meta-analysis. J Med Internet Res 2019 Jul 17;21(7):e11086 [FREE Full text] [doi: 10.2196/11086] [Medline: 31317869]
88. Moman RN, Dvorkin J, Pollard EM, Wanderman R, Murad MH, Warner D, et al. A systematic review and meta-analysis of unguided electronic and mobile health technologies for chronic pain-is it time to start prescribing electronic health applications? Pain Med 2019 Nov 01;20(11):2238-2255. [doi: 10.1093/pm/pnz164] [Medline: 31386151]
89. de Oliveira Lima L, Saragiotto BT, Costa LO, Nogueira LC, Meziat-Filho N, Reis FJ. Self-guided Web-based pain education for people with musculoskeletal pain: a systematic review and meta-analysis. Phys Ther 2021 Oct 01;101(10):pzab167. [doi: 10.1093/ptj/pzab167] [Medline: 34174081]
90. Takeshima N, Sozu T, Tajika A, Ogawa Y, Hayasaka Y, Furukawa TA. Which is more generalizable, powerful and interpretable in meta-analyses, mean difference or standardized mean difference? BMC Med Res Methodol 2014 Feb 21;14:30 [FREE Full text] [doi: 10.1186/1471-2288-14-30] [Medline: 24559167]
91. Deeks JJ, Higgins JP, Altman DG. Chapter 9: Analyzing data and undertaking meta-analyses. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. London, UK: Cochrane Collaboration; 2011.
92. Identifying publication bias in meta-analyses of continuous outcomes. Cochrane Training. 2020. URL: https://training.cochrane.org/resource/identifying-publication-bias-meta-analyses-continuous-outcomes [accessed 2022-06-24]

---

## Abbreviations

**PCA:** principal component analysis
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PROSPERO:** International Prospective Register of Systematic Reviews
**RCT:** randomized controlled trial
**RMIC:** Rainbow Model of Integrated Care
**SMD:** standardized mean difference

**WHO:** World Health Organization

---

*Edited by R Kukafka; submitted 10.03.22; peer-reviewed by A Rovetta; comments to author 14.04.22; revised version received 17.05.22; accepted 25.07.22; published 06.09.22.*

*Please cite as:*
*Valentijn PP, Tymchenko L, Jacobson T, Kromann J, Biermann CW, AlMoslemany MA, Arends RY*
*Digital Health Interventions for Musculoskeletal Pain Conditions: Systematic Review and Meta-analysis of Randomized Controlled Trials*
*J Med Internet Res 2022;24(9):e37869*
*URL: https://www.jmir.org/2022/9/e37869*
*doi:10.2196/37869*
*PMID:36066943*

Original Paper

# A Study of Publicly Available Resources Addressing Legal Data-Sharing Barriers: Systematic Assessment

Mohammad A Karim[1,2], PhD; Hye-Chung Kum[1,2], PhD; Cason D Schmit[1,2,3], JD

[1]Department of Health Policy and Management, School of Public Health, Texas A&M University, College Station, TX, United States
[2]Population Informatics Lab, School of Public Health, Texas A&M University, College Station, TX, United States
[3]Health Law and Policy Program, School of Public Health, Texas A&M University, College Station, TX, United States

**Corresponding Author:**
Cason D Schmit, JD
Department of Health Policy and Management
School of Public Health
Texas A&M University
1266 TAMU
212 Adriance Lab Rd
College Station, TX, 77843
United States
Phone: 1 979 436 0277
Email: schmit@tamu.edu

## Abstract

**Background:** United States data protection laws vary depending on the data type and its context. Data projects involving social determinants of health often concern different data protection laws, making them difficult to navigate.

**Objective:** We systematically aggregated and assessed useful online resources to help navigate the data-sharing landscape.

**Methods:** We included publicly available resources that discussed legal data-sharing issues with some health relevance and published between 2010 and 2019. We conducted an iterative search with a common string pattern using a general-purpose search engine that targeted 24 different sectors identified by Data Across Sectors for Health. We scored each online resource for its depth of legal and data-sharing discussions and value for addressing legal barriers.

**Results:** Out of 3710 total search hits, 2721 unique URLs were reviewed for scope, 322 received full-text review, and 154 were selected for final coding. Legal agreements, consent, and agency guidance were the most widely covered legal topics, with HIPAA (The Health Insurance Portability and Accountability Act), Family Educational Rights and Privacy Act (FERPA), Title 42 of the Code of Federal Regulations Part 2 being the top 3 federal laws discussed. Clinical health care was the most prominent sector with a mention in 73 resources.

**Conclusions:** This is the first systematic study of publicly available resources on legal data-sharing issues. We found existing gaps where resources covering certain laws or applications may be needed. The volume of resources we found is an indicator that real and perceived legal issues are a substantial barrier to efforts in leveraging data from different sectors to promote health.

***(J Med Internet Res 2022;24(9):e39333)*** doi:10.2196/39333

**KEYWORDS**

privacy; confidentiality; public health informatics; data sharing

## Introduction

Increasingly, data are leveraged to promote health outcomes, and practitioners are increasingly using data from different sectors to address social determinants of health. Unfortunately, the United States does not have a comprehensive data protection law; instead, there is a patchwork of laws that vary depending on the data type, who has it, and what they want to do with it [1-3]. Consequently, federal data protection laws vary considerably, and these differences are magnified by differences between state and local governments within the United States. The variation in data protection laws is particularly vexing for efforts in promoting data sharing to promote population health [4]. For example, the absence of a public health exception in the Title 45 of the Code of Federal Regulations (CFR) Part 2 protections for substance abuse treatment data has posed a major challenge to leveraging data to combat the opioid epidemic [5]. Commonly, practitioners are confronted with legal barriers to data use; some are real barriers (eg, legal language prohibiting data use), but many are perceived legal barriers (eg, perceptions that laws restrict data use) [2,6]. For example, the HIPAA (Health Insurance Portability and Accountability Act) rule has a robust public health data use exception [7], but it has been frequently cited as a (perceived) barrier to sharing data for public health purposes (eg, the 2013 fungal meningitis outbreak) [6]. Practitioners pursuing multisectoral data projects are forced to navigate the real and perceived legal barriers from the patchwork of US data protection laws [1,2,8].

Public health practitioners are specially affected because a combination of data on different aspects of a person's life and health may prove necessary to make the best-informed decisions. For example, education is a potent social determinant of health [9]. Consequently, there is substantial interest in determining whether laws (eg, the Family Educational Rights and Privacy Act [FERPA] and HIPAA) permit linking education attainment data with health outcomes data to further understand this social determinant of health [10-12]. Importantly, while leveraging data for public health has support among the US public, many data silos are often reinforced by legal restrictions. This sometimes leads to suboptimal cross-sector collaboration, ultimately resulting in less-than-ideal population health efforts. Nevertheless, navigating these different legal data protection frameworks is essential to achieving the goal of implementing precision public health because data on social determinants of health (eg, education, crime, and housing) will implicate several different data protection laws [8].

For example, previous studies have indicated the value of data linkage in identifying the association between health and health determining factors, such as income and crime [13,14]. Among recent initiatives, the efforts to link National Center for Health Statistics (NCHS) data and US Department of Housing and Urban Development (HUD) administrative records is an example of collaboration between 2 federal agencies that enabled linkage of housing and health data where the agencies used a memorandum of understanding to comply with relevant regulations [15]. Detailed guidance on addressing the legal challenges involved in these types of data sharing and linking efforts can provide useful reference points for practitioners at the state and local levels.

Publicly available online resources can help practitioners navigate these issues and inform conversations with legal counsel. Publicly available resources can help practitioners to understand whether laws exist that might protect certain data (eg, education, substance abuse treatment, juvenile justice, and government nutrition program data) [16]. However, without more detailed discussion, general descriptions of laws could beget perceived data-sharing barriers that could discourage pursuit of a proposed data-sharing project. However, laws that protect data often permit data to be used for secondary purposes [8,17]. Consequently, the most valuable and helpful publicly available resources on legal data-sharing issues contain detailed discussion of data protection laws, including both restrictions and permissions [18]. Detailed publicly available resources are becoming increasingly important as health informatics projects begin to span data sources in the effort to understand the social determinants of health.

In the field, publicly available resources are often the first resort (ie, Google searches). The presence or absence of quality resources describing legal mechanisms for data sharing can impact decisions to pursue data-sharing projects for public health purposes. Ideally, a resource goes beyond identifying legal issues and actually applies the law to specific use cases [18]. This type of use-case analysis can help public health professionals understand what is legally possible and help professionals identify relevant legal issues to discuss with their legal counsel. Although following professional legal advice is imperative for any data-sharing project, the existence of publicly available legal resources can be highly influential in the earliest planning stages and can sway leadership decisions on whether to pursue official legal counsel or abandon a project idea at inception.

However, finding quality resources discussing legal data-sharing issues can be challenging. For example, many documents discussing data sharing may make a passing reference to challenges posed by privacy laws and may even name a law (eg, HIPAA) [19-21]. However, quality (and helpful) legal analysis usually requires applying laws to facts using case studies or examples to show how the law operates in given situations [22,23]. Documents that only superficially reference privacy or legal data-sharing barriers are not helpful to practitioners and may even bury quality resources in search results.

There have been efforts by different organizations to facilitate data sharing across sectors, and many approaches have been documented. For example, in 2017, Data Across Sectors for Health (DASH) and the Network for Public Health Law developed the Legal Bibliography and more recently the DASH Knowledge Base, an online database of publicly available data-sharing resources to help public health practitioners navigate these complex legal issues [24]. This review is an extension of this work. However, these resources have not been systematically studied. Understanding this landscape is critical to understanding practitioners' current focus areas, specific challenges and needs, and what gaps exist in the existing literature.

This review focuses entirely on public resources (eg, white papers, reports, toolkits, and open-access academic articles) that are freely available to laypersons and practitioners. Prior reviews have explored data-sharing issues, but these are mostly academically focused (ie, sharing research data between academics) [25]. Reviews of resources concerned with combating legal barriers of data sharing in nonresearch settings are nonexistent in academic literature.

In this review, we have aggregated and screened through those publicly available resources that may help public health officials and practitioners navigate the data-sharing landscape. In recognition that health is affected by a tremendous number of factors (eg, social determinants of health) and assuming that future public health informatics application (eg, precision public health) can leverage these data for public health purposes, we were inclusive to the broadest extent in identifying resources that cover sectors affecting an individual's health.

## Methods

### Scope

We collected publicly available internet resources discussing data-sharing legal issues relevant to health. We used a broad interpretation of factors affecting health, considering any factor directly or indirectly affecting the well-being of an individual as a potential determinant of health. We only included resources if they were free to access and publicly available (including open-access academic articles). Academic articles were only included if they met our inclusion criteria and were freely accessible. We omitted results published prior to 2010 to ensure that the resources were reasonably current; laws change, and at least one major health-related data-sharing law, the Health Information Technology for Economic and Clinical Health Act, was enacted in 2009. We also excluded resources where the law or legal issue was not discussed with particularity; that is, resources that merely referenced a law or legal issue without some discussion were omitted. Some documents—like news articles, unannotated legislative text, and organizational policy statements—were excluded because they were not developed as "resources."

### Collection

We used a general-purpose search engine (Google) to identify the resources because a consumer-focused search engine is likely to be a common (if not default) search tool used by practitioners to learn about data-sharing issues.

To ensure a comprehensive search scope, we developed a complex search pattern yielding 75 individual searches, rather than using a single search term. Each search included a common string pattern: (common search stem) + (sector) + (data protection term). The common search stem applicable to each search was as follows: ("data sharing" OR "data use" OR "information sharing" OR "information use") + ("law" OR "regulation" OR "legal" OR "statute"). We identified search terms to target a total of 24 different sectors (sectors were identified in collaboration with the Data Across Sectors for Health and the Network for Public Health Law) [26], and an additional set of searches was executed without a specified sector (24 sectors and 1 overall). The common search stem and the sector search terms were executed a total of 3 times, each with a different data protection term: "privacy," "confidentiality," or "consent" (in that order). This search pattern yielded a total of 75 individual searches (ie, 25×3=75), and the first 50 hits for each search were saved. We justified capping our individual search results at 50 on the basis that individuals do not often view more than 5 pages of Google search results. The initial search was completed in September 2019.

### Coding

Two researchers (CS and MK) coded each resource independently. One researcher (CS) had a legal background and expertise in legal data-sharing issues, and the other researcher (MK) had a health services research background with expertise in data analysis. We used coding meetings to resolve discrepancies.

We scored each online resource on a scale of 1 to 4 (lowest 1, highest 4) in terms of their depth of legal issues discussed, depth of data-sharing discussion, and value for addressing legal barriers. We calculated interrater reliability scores for these 3 measures using Gwet's AC2 for ordinal data [27]. We used the objective benchmarking standards proposed by Altman [27] to interpret the AC2 coefficients (where a score of <0.20 represented "Poor", 0.21 to 0.40 represented "Fair", 0.41 to 0.60 represented "Moderate", 0.61 to 0.80 represented "Good", and 0.81 to 1.00 represented a "Very Good" strength of agreement) [27]. The calculated Gwet's AC2 scores for depth of legal discussion and overall value of resource were 0.59, indicating that agreement on these 2 measures approached the "Good" strength of agreement benchmark. The AC2 score for depth of data-sharing discussion was lower at 0.40, indicating that agreement on this measure approached the "Moderate" strength of agreement benchmark. The following section contains brief descriptions of our coding criteria for these 3 items. However, Multimedia Appendix 1 describes the coding criteria in greater detail.

Codes for the depth of legal discussion were primarily determined by the presence and extent of 2 factors: (1) discussion or description of the law or legal issues and (2) application of the law or legal issue on a specific set of facts (eg, à la tradition legal analysis). For example, a resource that contained both a detailed description of the law and applied the law to a specific use case would earn the highest score of 4 for legal depth. However, if a resource either described the law in detail or provided an extended discussion of how the law was applied in specific use cases, but did not do both, the score the resource received was lowered to 3 instead of 4. In contrast, a resource that identified the law or legal issue related to specific use cases and provided only basic information about the law earned a score of 2, whereas a resource that contained only a superficial description of the law or legal issue earned the lowest score of 1. In addition to these criteria, we had another criterion for template legal agreements. Template agreements with extensive annotations (ie, explaining the purpose or function of contractual terms) were coded with the highest legal depth (score of 4), and template agreements with moderate or without annotations were scored lower (score of 3 or 2, respectively). Importantly, the coding of the depth of legal discussion did not consider the quality or legal accuracy of the discussion nor whether the discussion appears consistent with the referenced statutes, regulations, or related judicial interpretations.

The coding on the depth of data-sharing discussion evaluated the extent the resource covers strategies to initiate or maintain at least 1 type of data-sharing activity. Codes for the depth of data-sharing discussion were primarily determined by the presence and extent of 2 factors: (1) discussion or description of a data-sharing issue and (2) discussion or description of a data-sharing strategy or process. Two additional factors separated the highest-scoring resources on data sharing: (1) use cases explaining data-sharing issues and strategies in specific contexts and (2) links to recommended additional data-sharing resources.

Finally, we assessed the value of each resource for addressing legal barriers to nonexpert users. The codes on the overall value of the resource for addressing legal barriers were based on the presence of several factors. Some factors weighed in favor of higher scores, including if the resource was highly scored for legal or data-sharing discussion, user-friendly, or from an official governmental source. Other factors weighed against a higher score, including if the resource contained only limited context (eg, PowerPoint slides) or if the relevant discussion was only tangential to the focus of the resource (eg, a resource that includes an overview of legal or data-sharing issues as an appendix to the main document).

## Results

### Overview

Our sector-specific searches provided a total of 3710 hits, out of which 989 were duplicates. After removing the duplicates, the remaining 2721 unique URLs were subjected to scoping screening. The full text of 322 in-scope resources were reviewed, out of which 154 were selected for final coding (Figure 1). Table S1 in Multimedia Appendix 2 includes a list of all included resources, their sectors covered, and their scores for legal depth, data-sharing depth, and value.

Common resources excluded were company privacy statements, commentaries or analysis [28], slideshow documents with superficial information [29], and sources that have no discussion of US law [30,31]. Among the resources selected for coding, an upward trend in number between the years 2010 and 2018 was observed (Figure 2).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) chart showing the scoping process for collected records and publicly available resources. doc: document; Wiki: Wikipedia.

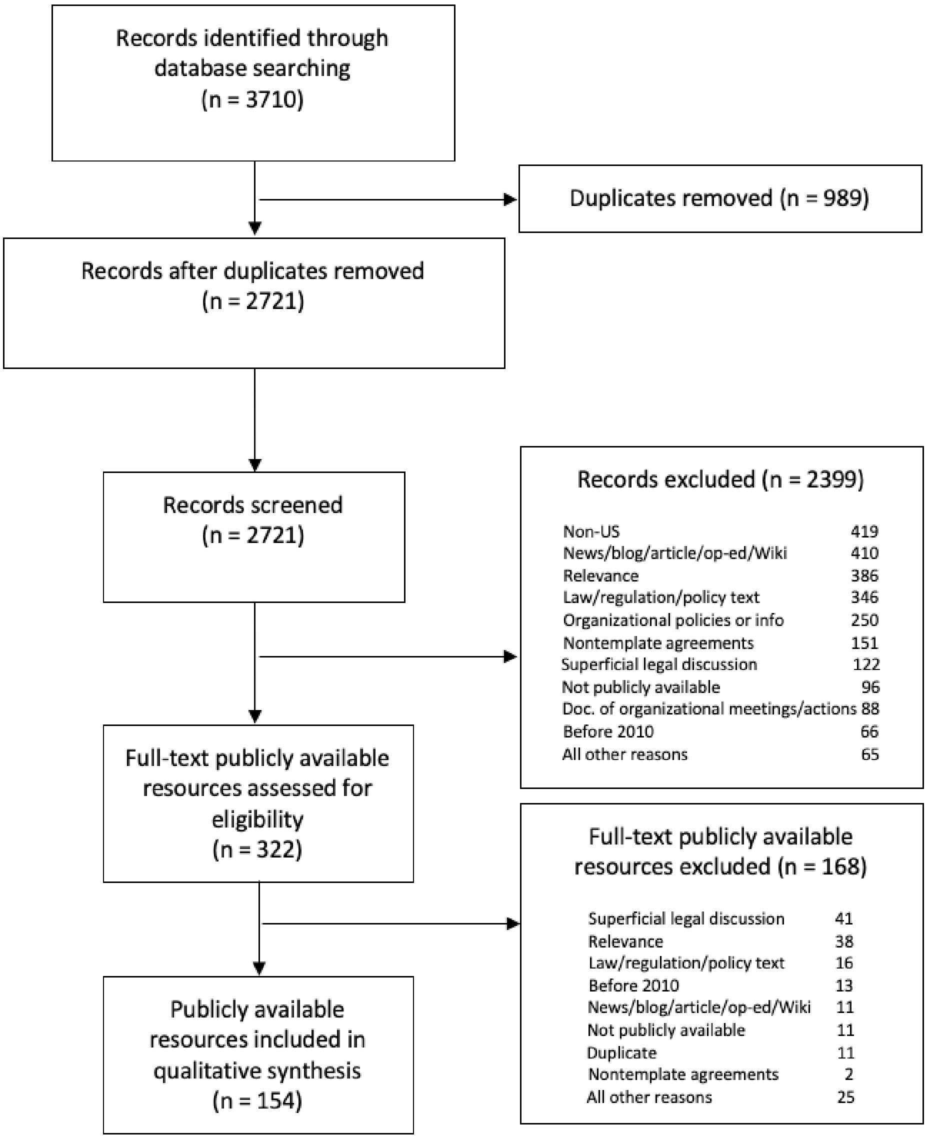
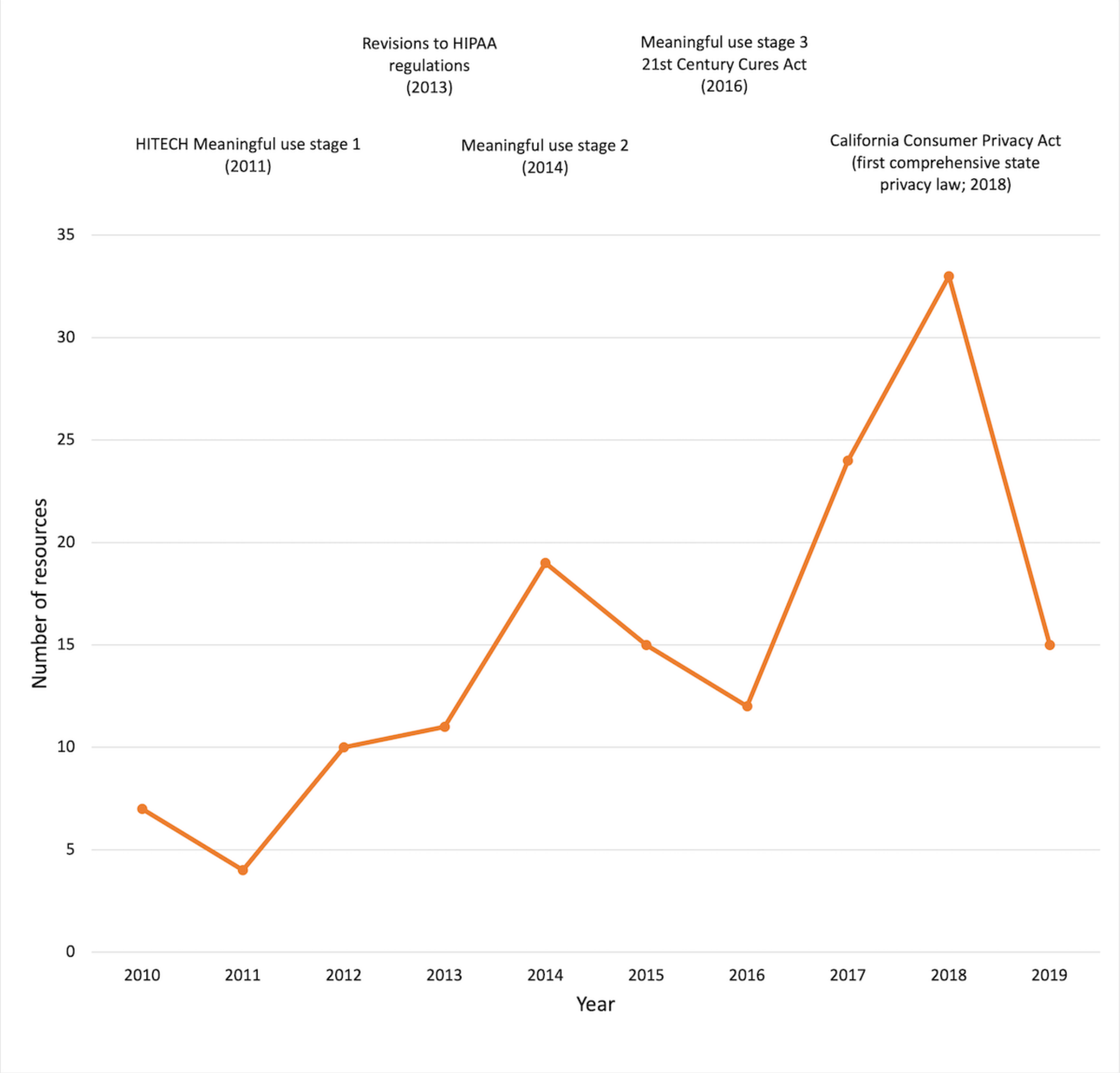


**Figure 2.** Number of publicly available resources identified per year and significant legal developments in privacy. Note that the 2019 data represent a partial year (January to September) with average monthly resources dropping from 2.75 in 2018 to 1.67 in 2019. Four resources identified in our 2019 search were subsequently updated in 2020 prior to coding completion. HITECH: Health Information Technology for Economic and Clinical Health Act. HIPAA: The Health Insurance Portability and Accountability Act.



## Legal Topics and Resources Covered in Resources

Legal agreements for data sharing (n=63) were the most commonly available resource, followed by consent (n=57) and agency guidance (n=57; Table 1). Among those legal topics covered in at least 10 resources, legal case studies had the highest mean scores in terms of legal depth (mean score 3.38, SD 0.75), data-sharing depth (mean score 3.12, SD 1.11), and resource value (mean score 3.31, SD 0.79).

**Table 1.** Legal subject or topic covered by publicly available resources, mean depth of legal discussion and data-sharing discussion, and mean resource value (N= 154).

| Legal resource | Resources present in, n | Legal depth score, mean (SD) | Data-sharing depth score, mean (SD) | Resource value score, mean (SD) |
|---|---|---|---|---|
| Legal agreements for data sharing | 63 | 2.90 (0.91) | 2.57 (1.10) | 2.94 (0.86) |
| Consent (obtaining consent, waiving requirements, models) | 57 | 3.12 (0.85) | 2.68 (1.07) | 3.05 (0.89) |
| Agency resource or guidance | 57 | 3.00 (0.85) | 2.12 (1.13) | 2.81 (0.93) |
| Interagency data sharing | 50 | 3.14 (0.90) | 2.86 (1.09) | 3.00 (0.97) |
| General legal overview (ie, no specific application indicated) | 41 | 2.88 (0.93) | 1.98 (1.06) | 2.51 (0.98) |
| Resource links | 34 | 3.09 (0.83) | 2.79 (1.12) | 3.06 (0.81) |
| Case studies applying law | 26 | 3.38 (0.75) | 3.12 (1.11) | 3.31 (0.79) |
| Provider sharing | 22 | 3.00 (0.87) | 2.82 (1.01) | 2.91 (1.02) |
| Health authority use | 22 | 3.18 (0.85) | 2.82 (0.96) | 3.09 (1.06) |
| Health information exchange | 19 | 3.00 (0.94) | 2.63 (1.07) | 2.84 (1.01) |
| Frequently asked questions about law | 14 | 3.21 (0.70) | 2.71 (1.27) | 2.93 (1.00) |
| Other legal resource | 13 | 2.92 (1.04) | 2.69 (1.18) | 2.85 (1.34) |
| Data system governance | 8 | 2.75 (0.71) | 3.50 (0.53) | 3.00 (0.93) |
| Data sharing for program evaluation | 7 | 3.57 (0.53) | 3.57 (0.79) | 3.71 (0.49) |
| Court orders/subpoenas | 4 | 3.25 (0.50) | 3.25 (0.96) | 3.50 (0.58) |
| Working with legal counsel | 4 | 2.25 (0.96) | 3.25 (0.96) | 3.00 (0.82) |
| Medical-legal partnerships | 2 | 3.50 (0.71) | 3.50 (0.71) | 3.50 (0.71) |
| Model legislation | 2 | 2.00 (0.00) | 1.50 (0.71) | 1.50 (0.71) |
| Statistical methods for protecting privacy and confidentiality | 1 | 2 (N/A[a]) | 4 (N/A) | 3 (N/A) |

[a] N/A: not applicable.

## Laws Discussed in Resources

The reviewed resources covered a total of 96 laws or legal issues to different degrees, but the plurality of them focused on only a handful of laws. Out of the 97 laws discussed in total, only 16 were discussed in at least 4 resources (Table 2). HIPAA (n=74), FERPA (n=41), and 42 CFR Part 2 (n=35) were the top 3 federal laws discussed in the resources [32,33]. See Table S2 Multimedia Appendix 2 for laws discussed in fewer than 4 resources. Among the laws discussed in at least 10 resources, the Privacy Act of 1974 scored the highest mean scores in terms of legal depth (mean score 3.45, SD 0.69), and FERPA scored the highest on data-sharing depth (mean score 2.95, SD 1.09) and resource value (mean score 3.22, SD 0.88). Among the 154 full-text resources coded, 68.8% (n=106) discussed more than 1 law (Supplemental Figure S1, Multimedia Appendix 2).

**Table 2.** Laws discussed in publicly available resources, mean depth of legal discussion and data-sharing discussion, and mean resource value (N=154).

| Law | Resources present in, n | Legal depth score, mean (SD) | Data-sharing depth score, mean (SD) | Resource value score, mean (SD) |
|---|---|---|---|---|
| Health Insurance Portability and Accountability Act | 74 | 3.01 (0.88) | 2.65 (1.13) | 2.93 (0.94) |
| Other state or local law(s) | 43 | 2.91 (0.92) | 2.37 (1.09) | 2.79 (0.94) |
| Family Educational Rights and Privacy Act | 41 | 3.24 (0.83) | 2.95 (1.09) | 3.22 (0.88) |
| 42 CFR[a] Part 2 | 35 | 3.26 (0.82) | 2.91 (1.07) | 3.09 (0.95) |
| General legal concepts | 22 | 2.27 (1.03) | 2.18 (1.10) | 2.45 (1.06) |
| The Privacy Act of 1974 | 11 | 3.45 (0.69) | 2.64 (1.43) | 3.18 (0.98) |
| Health Information Technology for Economic and Clinical Health Act | 10 | 3.10 (0.99) | 2.80 (1.03) | 2.80 (0.92) |
| Freedom of Information Act (or similar state laws) | 9 | 3.44 (0.53) | 2.56 (1.33) | 3.00 (1.12) |
| Medicaid privacy requirements | 8 | 2.62 (0.92) | 2.38 (0.92) | 2.75 (1.04) |
| Federal Policy for the Protection of Human Subjects (Common Rule) | 8 | 2.38 (0.92) | 2.12 (1.13) | 2.25 (0.89) |
| Workforce Innovation and Opportunity Act | 7 | 3.71 (0.49) | 2.71 (1.11) | 3.43 (0.79) |
| Confidentiality protections governing unemployment compensation wage records | 5 | 3.40 (0.55) | 2.80 (1.30) | 3.40 (0.89) |
| The McKinney-Vento Homeless Assistance Act of 1987 | 5 | 2.60 (0.89) | 2.80 (1.10) | 2.80 (0.84) |
| Individuals with Disabilities Education Improvement Act | 4 | 3.25 (0.96) | 3.25 (0.96) | 3.25 (0.96) |
| Confidential Information Protection and Statistical Efficiency Act of 2002 | 4 | 3.50 (0.58) | 3.25 (0.96) | 3.50 (0.58) |
| Food Stamp Act of 1964 | 4 | 2.75 (0.50) | 2.50 (1.00) | 2.75 (0.50) |

[a] CFR: Code of Federal Regulations.

## Data Use Cases Covered in Resources

The most frequently addressed use case was record matching across systems with 113 resources discussing this (Table 3). Statistical analysis was the second most discussed use case (n=58). This was followed by reporting function, a use case that was discussed in 44 resources. Among the data use cases referenced in at least 10 resources, calculating and reporting metrics scored the highest for legal depth (mean score 3.40, SD 0.71) and resource value (mean score 3.36, SD 0.70), while generating predictive scores scored the highest on data-sharing depth (mean score 3.00, SD 1.20).

**Table 3.** Data use case discussed in publicly available resource, mean depth of legal discussion and data-sharing discussion, and mean resource value (N=154).

| Use case | Resources present in, n | Legal depth score, mean (SD) | Data-sharing depth score, mean (SD) | Resource value score, mean (SD) |
|---|---|---|---|---|
| Using identifying information to match records across systems to create a more encompassing view of a person or case | 113 | 2.90 (0.97) | 2.62 (1.12) | 2.85 (0.99) |
| Statistical analysis to look for useful patterns and relationships in the data set | 58 | 3.12 (0.86) | 2.79 (1.10) | 3.09 (0.88) |
| Reporting functions that allow users to specify and generate reports using items from a menu | 44 | 3.11 (0.84) | 2.45 (1.21) | 3.05 (0.86) |
| Calculating and reporting of metrics, indicators, and dashboards enabling group comparison and tracking of progress over time | 25 | 3.40 (0.71) | 2.68 (1.31) | 3.36 (0.70) |
| Generating scores that predict/identify likelihood or risk of future events | 22 | 3.14 (0.77) | 3.00 (1.20) | 3.23 (0.87) |
| Automating decision support and generating recommendations or alerts | 21 | 3.14 (0.91) | 2.38 (1.28) | 2.95 (0.92) |
| Not expressly discussed in resource | 14 | 2.64 (0.93) | 1.79 (0.97) | 2.50 (0.94) |
| Presentation and visualization of data such that the viewer grasps the relevance of the information | 10 | 2.70 (0.82) | 2.00 (1.25) | 2.60 (0.97) |
| Other use case | 8 | 2.88 (0.99) | 1.88 (1.36) | 2.88 (0.83) |
| Mapping/geographic information systems—analysis of data by geographic location and presentation as maps | 4 | 3.25 (0.96) | 2.75 (1.50) | 3.50 (1.00) |

## Sectors Covered in Resources

A total of 20 sectors were covered as the primary focus of the included resources, among which education (n=22), public health (n=16), and academia (n=13) were most common. Among these 3, education was the highest-scoring sector in terms of legal depth (mean score 3.09, SD 0.92), data-sharing depth (mean score 2.82, SD 1.22), and resource value (mean score 3.05, SD 0.95). The 4 sectors that were initially searched but not considered as a representative sector for any of the resources in the final data set were “elected or appointed official,” “faith or faith based,” “parks and recreation,” and “philanthropy.” Among the sectors, clinical health care was the most prominent sector with a mention in 73 websites or files (Table 4). Among the 154 resources, around 86.4% (n=133) discussed more than 1 sector (Figure S1, Multimedia Appendix 2).

A relatively small proportion of our 2721 unique search results were scored as having the greatest depth of legal discussion (n=48, 1.76% among unique search results) [32,33] and depth of data-sharing discussion (n=36, 1.32% among unique search result). Instead, most search results were either international (eg, did not address US laws) [30], out-of-date, provided only legislative updates [34], had a specific focus unrelated to health, or only contained passing or superficial discussion of legal data-sharing issues. Additionally, we found resources that were blog posts [35], PowerPoint slides with very limited information, privacy statements on commercial sites [36], policy memoranda [37], or organization-specific policies [38], or that only defined a law without providing any further discussion [39].

**Table 4.** Sectors addressed as the main or primary focus of a resource, number of resources addressing the sector as a secondary focus, mean depth of legal discussion and data-sharing discussion, and mean resource value (N=154).

| Sector | Primary focus | | | | Secondary focus |
|---|---|---|---|---|---|
| | Resources present in, n | Legal depth score, mean (SD) | Data-sharing depth score, mean (SD) | Resource value score, mean (SD) | Resources present in, n |
| Education/schools | 22 | 3.09 (0.92) | 2.82 (1.22) | 3.05 (0.95) | 45 |
| Public health (government) | 16 | 2.94 (1.06) | 2.62 (1.09) | 2.88 (1.15) | 30 |
| Academia/research | 13 | 2.38 (1.04) | 2.08 (1.04) | 2.54 (1.05) | 23 |
| Clinical health care | 11 | 2.91 (0.94) | 1.82 (0.87) | 2.64 (0.92) | 73 |
| Social and human services | 11 | 3.27 (0.79) | 2.64 (1.03) | 3.09 (0.83) | 27 |
| Multiple sectors | 11 | 2.91 (0.94) | 2.55 (1.04) | 3.00 (1.00) | N/A[a] |
| Organized government (tribal/local/state/federal) not included in others | 10 | 2.90 (1.20) | 2.70 (1.25) | 3.00 (1.05) | 33 |
| Mental/behavioral health care | 10 | 3.00 (0.94) | 2.50 (1.18) | 2.70 (1.06) | 33 |
| Information management infrastructure | 9 | 2.78 (1.20) | 3.11 (1.05) | 2.78 (1.09) | 22 |
| Public safety/law enforcement | 7 | 3.43 (0.53) | 1.71 (1.11) | 2.43 (1.13) | 22 |
| Housing and homelessness | 6 | 2.33 (0.52) | 1.83 (0.98) | 2.50 (0.55) | 11 |
| Business | 5 | 3.20 (0.45) | 2.20 (1.30) | 3.40 (0.55) | 17 |
| Health care payers | 4 | 2.00 (0.82) | 1.75 (0.96) | 1.75 (0.96) | 48 |
| Criminal justice/correctional facilities | 4 | 3.00 (0.82) | 2.75 (1.26) | 3.00 (0.82) | 10 |
| Justice system/courts | 4 | 2.50 (0.58) | 2.25 (0.50) | 2.50 (0.58) | 17 |
| Food and nutrition | 3 | 3.33 (0.58) | 2.67 (1.53) | 3.00 (1.00) | 5 |
| Banking/financial | 3 | 2.00 (0.00) | 2.33 (1.53) | 2.00 (0.00) | 7 |
| Legal/law firms | 1 | 4 (N/A) | 4 (N/A) | 4 (N/A) | 3 |
| Not expressly discussed in resource | 1 | 2 (N/A) | 4 (N/A) | 3 (N/A) | 1 |
| Other community-based, community action group | 1 | 4 (N/A) | 3 (N/A) | 3 (N/A) | 2 |
| Planning, economic, or community development | 1 | 2 (N/A) | 4 (N/A) | 2 (N/A) | 1 |
| Transportation/infrastructure | 1 | 3 (N/A) | 4 (N/A) | 4 (N/A) | 3 |
| Other | N/A | N/A | N/A | N/A | 11 |

[a] N/A: not applicable.

## Discussion

This is the first systematic study of publicly available resources on legal data-sharing issues. Publicly available resources are often the resources of first resort for practitioners, and the presence or absence of resources may factor in decisions to pursue a data-sharing project or engage with legal counsel. Consequently, it is important to understand what resources exist and what gaps are present. This paper helps map the existing landscape and can inform future work. For example, a number of quality resources exist for laws that govern health data, but fewer resources exist that discuss legal data-sharing issues pertaining to other social determinants of health, such as housing and homelessness.

It is possible that high numbers of resources addressing the same law might be an indicator of legal complexity or perceived legal barriers associated with that law. For example, HIPAA was one of the laws that was discussed in the most resources; however, HIPAA has generous exceptions that permit using data for public health and research purposes [6]. The fact that so many resources address HIPAA as a legal issue facing data sharing could be an indicator that the law is overly complex, misunderstood, or conservatively applied by organizations. Alternatively, the presence of a large number of publicly available resources addressing a law could indicate the law's importance or significance to the activities of data custodians or simply a greater demand for knowledge and awareness of the law or legal issue.

Our findings suggest that good resources are difficult to find. Practitioners trying to find pertinent resources will have to sift through voluminous search results that are not useful to identify, understand, and address legal barriers to data sharing. Consequently, the difficulty of finding quality resources likely amplifies the perception of legal data-sharing barriers among practitioners.

To our surprise, the resources we identified cite nearly 100 different federal data-sharing laws or legal issues. The number of laws and legal issues was far higher than we expected. It suggests data-sharing challenges extend far beyond HIPAA, FERPA, and 42 CFR Part 2. Moreover, we also identified a large number of resources addressing multiple data protection laws or multiple sectors. These findings suggest that practitioners are working to address cross-sectoral legal data-sharing challenges. Given the patchwork legal data protection framework that exists in the United States, data-sharing projects designed to address the social determinants of health will likely cross multiple sectors and implicate the different data protection laws associated with those different sectors. The data silos—reinforced by these different data protection laws—have been cited as a barrier to the study of social determinants of health [2,4]. Our findings suggest that addressing these cross-sectoral challenges could be driving the development of publicly available legal resources. These challenges could be addressed with a comprehensive data-sharing framework [2,8]. For example, the European General Data Protection Regulation provides a straightforward legal analysis for cross-sectoral data sharing because it provides a common set of legal definitions, rules, and exceptions for all data controllers and custodians; in contrast, the United States has up to 6 different privacy laws that could apply to veterans' health information [8,40]. Although privacy scholars have long cited the need for a comprehensive privacy law in the United States, Congress has struggled to appease the broad and diverse stakeholders for a national privacy law [8,41].

Our findings also suggest that many publicly available documents focus on legal agreements and consent documents that enable data sharing. Developing legal agreements from scratch can be incredibly expensive given the cost of legal services. Good template agreements can reduce costs tremendously and can be valuable starting points for legal counsel. Given these considerations, it is understandable that so many publicly available resources would address legal data-sharing agreements and consent documents. However, the template agreements we found varied in quality and utility. For example, some template agreements contained annotations that explained the purpose or function of specific terms and provisions [42,43], while some did not [44-46]. These annotations are useful to ensuring that agreements are well tailored to the needs of the data project. Without these annotations, there is a risk of contracting parties relying on a sample agreement that can inadvertently include counterproductive terms as boilerplate language.

We also note that we identified an interesting trend in publications over time. Our data show an increase in the number of publicly available resources relating to legal data-sharing issues from 2010 (ie, the earliest date within our scope) to 2018 but then a sudden decrease in 2019. It is possible that the increase in publications could be driven by the implementation of new data-sharing legislation and related regulations—like the federal Health Information Technology for Economic and Clinical Health Act (HITECH) Act (2009) and the 21st Century Cures Act (2016)—which created new data-sharing legal tools and opportunities. Some of the observed decrease after 2018 could be because our September 2019 search did not include October, November, and December 2019 publications; however, this likely does not explain the drop in the publication rate between 2018 and 2019 (from 2.75 publications per month to 1.67 publications per month). The decrease could be explained by an increase in interest in the newly implemented General Data Protection Regulation (GDPR) in the European Union, which affected many privacy policies of domestic entities and organizations due to its broad scope. Any resources that solely addressed GDPR legal issues would have been excluded from this research as an international law. If organizations that create publicly available resources on privacy shifted focus—and their finite resources—to the GDPR after its implementation, this could explain the sudden drop in publications that we observed.

Although publicly available resources can be very useful to practitioners in overcoming data-sharing barriers, there are several limitations associated with these resources. A Google search of "privacy" yields several trillion results, but only the top 400 or so are viewable under Google's propriety platform. This is one example of the limitations inherent to systematically searching for publicly available resources using a propriety—and nontransparent—system. We took efforts to ensure that we sampled a broad range of this space, but these finding cannot be considered comprehensive. Moreover, publicly available resources are not necessarily permanent. Some highly rated resources identified in our search were later found to be unavailable in their original online locations [47], and quality resources, previously known to the authors, were not identified in this search [18]. A small number of our identified resources were updated prior to our completion of coding (ie, in 2020). Additionally, organizations may move or remove online resources, and we found this to be a common issue during our study. Thus, it can be difficult for practitioners to maintain an existing list of online resources. This highlights the need to develop a comprehensive and dynamic knowledge base that will compile and maintain publicly available data-sharing resources. For example, the resources identified in this study are now incorporated into the DASH Knowledge Base, an online, practitioner-focused database of data-sharing resources that includes tools to search for relevant and useful resources [24].

Additionally, our efforts to broadly sample this space (ie, through 75 separate searches) might have introduced some bias. For example, one of our search terms was "consent," which might have inflated the number of template data-sharing forms that we found.

Finally, we note that this search does not include resources published in 2020, 2021 or 2022, so our results do not include the temporary emergency actions impelled by the COVID-19 response. Although this is a limitation, we note that the pace of federal legislation in data protection is glacial [8]. To our knowledge, no federal data protection act has been passed by Congress since the 21st Century Cures Act in 2016 (although the 2020 Coronavirus Aid, Relief, and Economic Security [CARES] Act included funding for loosely defined data modernization efforts).

This is the first systematic study of publicly available resources on legal data-sharing issues. Our findings describe the existing landscape of publicly available resources addressing legal data-sharing issues and can help identify future needs. We found existing gaps—like the Juvenile Justice and Delinquency Prevention Act which was discussed in only 1 resource or medical-legal partnerships, which was discussed in only 2 resources—where a lack of existing resources covering certain laws or applications allows existing data-sharing uncertainties to persist. We also found existing areas of saturation where certain laws and applications are covered extensively (eg, HIPAA and FERPA), such that new resource development might prove wasteful or perhaps even bury high quality resources deeper in search results. Moreover, many resources that we identified addressed multiple sectors or data protection laws, possibly indicating that cross-sectoral data sharing is a current priority in health informatics. Nevertheless, the volume of resources we found is an indicator that real and perceived legal issues are a substantial barrier to efforts to leverage data from different sectors to promote health. Although many resources exist to help practitioners navigate these legal issues, good resources may be hard to find.

## Acknowledgments

This work was supported by Data Across Sectors for Health and the Network for Public Health Law. The study funders provided limited input on study design (ie, sector categories) but otherwise had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; or decision to submit the manuscript for publication. No other financial disclosures were reported by authors.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Protocol.
[DOCX File , 32 KB - jmir_v24i9e39333_app1.docx ]

Multimedia Appendix 2
Supplementary tables.
[DOCX File , 85 KB - jmir_v24i9e39333_app2.docx ]

## References

1. Hulkower R, Penn M, Schmit C. Privacy and confidentiality of public health information. In: Magnuson J, Dixon B, editors. Public Health Informatics and Information Systems. Cham, Switzerland: Springer; 2020:147-166.
2. Schmit C, Kelly K, Bernstein J. Cross sector data sharing: necessity, challenge, and hope. J Law Med Ethics 2019 Jun;47(2_suppl):83-86. [doi: 10.1177/1073110519857325] [Medline: 31298141]
3. Schmit C, Sunshine G, Pepin D, Ramanathan T, Menon A, Penn M. Transitioning from paper to digital: state statutory and regulatory frameworks for health information technology. Public Health Rep 2017;132(5):585-592 [FREE Full text] [doi: 10.1177/0033354917722994] [Medline: 28813620]
4. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129 Suppl 2:19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
5. The President’s commission on combating drug addiction and the opioid crisis October 20, 2017 meeting minutes. 2017. The Trump Whitehouse Archive - National Archives. 2017. URL: https://trumpwhitehouse.archives.gov/sites/whitehouse.gov/files/images/Minutes%20from%2010-20-17%20meeting.pdf [accessed 2022-01-19]
6. Improving your access to electronic health records during outbreaks of healthcare-associated infections. Association for State and Territorial Health Officials. 2016. URL: https://web.archive.org/web/20201027191124/https://www.astho.org/Toolkit/Improving-Access-to-EHRs-During-Outbreaks/Address-Patient-Privacy-Concerns/ [accessed 2022-07-15]
7. 45 CFR § 164.512. Cornell Law School. URL: https://www.law.cornell.edu/cfr/text/45/164.512 [accessed 2021-12-20]
8. Schmit C, Larson B, Kum HC. Data Privacy in the time of plague. SSRN Journal 2022;21(1):152-227. [doi: 10.2139/ssrn.3968130]
9. Galea S, Tracy M, Hoggatt KJ, Dimaggio C, Karpati A. Estimated deaths attributable to social factors in the United States. Am J Public Health 2011 Aug;101(8):1456-1465. [doi: 10.2105/AJPH.2010.300086] [Medline: 21680937]
10. Comparison of FERPA and HIPAA privacy rule for accessing student health data. Association of State Territorial Health Officials. 2011. URL: https://www.astho.org/programs/preparedness/public-health-emergency-law/public-health-and-schools-toolkit/comparison-of-ferpa-and-hipaa-privacy-rule/ [accessed 2021-12-20]
11. Information sharing: case examples. Education Development Center. 2013. URL: http://informationsharing.promoteprevent.org/ [accessed 2021-12-20]

12. Ohio Department of Education. Data sharing for program evaluation. 2018. URL: https://education.ohio.gov/getattachment/Topics/Research-Evaluation-and-Advanced-Analytics/Sections/Our-Work-Includes/Data_Sharing_Guidance_for_Districts.pdf.aspx?lang=en-US [accessed 2021-12-20]
13. Larney S, Burns L. Evaluating health outcomes of criminal justice populations using record linkage: the importance of aliases. Eval Rev 2011 Apr;35(2):118-128. [doi: 10.1177/0193841X11401695] [Medline: 21398273]
14. Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, et al. The association between income and life expectancy in the United States, 2001-2014. JAMA 2016 Apr 26;315(16):1750-1766 [FREE Full text] [doi: 10.1001/jama.2016.4226] [Medline: 27063997]
15. Sperling J, Lloyd P, Helms V, Brittain J. Administrative, legal, and organizational hurdles in data linkage: experiences and lessons from the recent National Center for Health Statistics (NCHS) - U.S. Department of Housing and Urban Development (HUD) data linkage project. Int J Popul Data Sci 2017 Apr 19;1(1):1. [doi: 10.23889/ijpds.v1i1.351]
16. Summary of selected federal laws and regulations addressing confidentiality, privacy and security. Office of the National Coordinator for Health Information Technology. 2010. URL: https://www.healthit.gov/sites/default/files/privacy-security/federal-privacy-laws-table2-26-10-final.pdf [accessed 2021-12-17]
17. Ramanathan T, Schmit C, Menon A, Fox C. The role of law in supporting secondary uses of electronic health information. J Law Med Ethics 2015;43 Suppl 1:48-51 [FREE Full text] [doi: 10.1111/jlme.12215] [Medline: 25846164]
18. Sharfstein JM, Chrysler D, Bernstein J, Armijos L, Tolosa-Leiva L, Taylor H, et al. Using electronic health data for community health: example cases and legal analysis. deBeaumont Foundation. 2017. URL: https://debeaumont.org/wp-content/uploads/2019/04/Electronic-Health-Data-Report-1.pdf [accessed 2022-08-20]
19. Health information exchange policy issues. Agency for Healthcare Research and Quality. 2021. URL: https://digital.ahrq.gov/key-topics/health-information-exchange-policy-issues [accessed 2021-10-26]
20. Data sharing, care coordination, and population health: a learning guide. AcademyHealth. 2017. URL: https://academyhealth.org/publications/2017-06/data-sharing-care-coordination-and-population-health-learning-guide [accessed 2021-12-03]
21. Negotiating one of the largest HIE agreements in history. Healthcare Information and Management Systems Society. 2016. URL: https://www.himssconference.org/sites/himssconference/files/pdf/173.pdf [accessed 2021-10-26]
22. Authorities and limitations in sharing information between public health agencies and law enforcement. Association of State and Territorial Health Officials. 2012. URL: https://www.astho.org/programs/preparedness/public-health-emergency-law/public-health-and-information-sharing-toolkit/authorities-and-limitations/ [accessed 2021-12-03]
23. Data-sharing tool kit for communities: how to leverage community relationships while protecting student privacy. US Department of Education. 2016. URL: https://www2.ed.gov/programs/promiseneighborhoods/datasharingtool.pdf [accessed 2021-12-03]
24. DASH Knowledgebase. Data Across Sectors for Health. 2021. URL: https://dash.asemiocamp.com/ [accessed 2022-08-15]
25. Rowhani-Farid A, Allen M, Barnett AG. What incentives increase data sharing in health and medical research? A systematic review. Res Integr Peer Rev 2017;2:4 [FREE Full text] [doi: 10.1186/s41073-017-0028-9] [Medline: 29451561]
26. Common data model. Data Across Sectors for Health. 2017. URL: https://dashconnect.org/allin [accessed 2018-03-07]
27. Gwet KL. Handbook of Inter-rater Reliability: The Definitive Guide To Measuring the Extent of Agreement Among Raters. Gaithersburg, MD, USA: Advanced Analytics, LLC; 2014.
28. Weiss NE. Legislation to facilitate cybersecurity information sharingconomic analysis. Congressional Research Service. 2015. URL: https://sgp.fas.org/crs/misc/R43821.pdf [accessed 2021-12-20]
29. North Carolina State University Center for Geospatial Analysis. Powering Energy Efficiency & Impacts Framework Project. 2018. URL: https://www.energy.gov/sites/prod/files/2018/08/f54/PEEIF-webinar.pdf [accessed 2021-12-20]
30. Telematics data sharing, competition law and privacy rights. Pinsent Masons. 2014. URL: https://www.pinsentmasons.com/out-law/analysis/telematics-data-sharing-competition-law-and-privacy-rights [accessed 2021-12-10]
31. Fukushima M, Austin C, Sato N, Maruyama T, Navarro E, Rocca M, et al. The Global academic research organization network: Data sharing to cure diseases and enable learning health systems. Learn Health Syst 2019 Jan;3(1):e10073 [FREE Full text] [doi: 10.1002/lrh2.10073] [Medline: 31245596]
32. Overview of the Privacy Act of 1974. US Department of Justice. 2015. URL: https://www.justice.gov/archives/opcl/conditions-disclosure-third-parties [accessed 2021-12-03]
33. Dear colleague letter: foster youth information-sharing. California Department of Education, California Department of Social Services, California Department of Justice. 2016. URL: https://oag.ca.gov/sites/all/files/agweb/pdfs/bcj/fy-info.pdf [accessed 2021-12-03]
34. Legislative update: revenue cap bill passes Senate at 3.5 percent, House Committee follows suit. Texas Municipal League. 2019. URL: https://www.tml.org/DocumentCenter/View/1151/LU2019-17 [accessed 2021-12-03]
35. Meadows A. To share or not to share? That is the (research data) question. Scholarly Kitchen. 2014. URL: https://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question/ [accessed 2021-12-03]
36. Uber privacy notice. Uber Technologies Inc. 2021. URL: https://www.uber.com/legal/en/document/ [accessed 2021-12-03]
37. Policy memo: nuclear security and information sharing. The Stanley Foundation. 2015. URL: https://stanleycenter.org/publications/policy_memo/SPCPMNuclearSecurity1115.pdf [accessed 2021-12-03]

38. Geospatial data access, distribution and use. Oakland county - Michigan. 2014. URL: https://www.oakgov.com/it/gis/Documents/GISDataPolicies.pdf [accessed 2021-12-03]
39. Information sharing. Office of Tribal Justice - Department of Justice. 2014. URL: https://www.justice.gov/otj/information-sharing [accessed 2021-12-03]
40. VHA Directive 1605.01: privacy and release of information. Department of Veteran Affairs. 2016. URL: https://dash.asemiocamp.com/ [accessed 2022-07-13]
41. Lima C. Top Senate Democrat casts doubt on prospect of major data privacy bill. The Washington Post. 2022 Jun 22. URL: https://www.washingtonpost.com/technology/2022/06/22/privacy-bill-maria-cantwell-congress/ [accessed 2022-07-13]
42. Instructions for completing the data use agreement (DUA) form CMS-R-0235. Centers for Medicare and Medicaid Services. 2010. URL: https://www.cms.gov/medicare/cms-forms/cms-forms/downloads/cms-r-0235.pdf [accessed 2021-12-03]
43. Submit a data sharing application. Military Health System. 2019. URL: https://www.health.mil/Military-Health-Topics/Privacy-and-Civil-Liberties/Submit-a-Data-Sharing-Application [accessed 2021-12-03]
44. P1 LERMS agency agreement. North Dakota Criminal Justice Information Sharing Division. 2019. URL: https://attorneygeneral.nd.gov/sites/ag/files/documents/CJISLERMSAgreement.pdf [accessed 2021-12-03]
45. Interagency HMIS Data Sharing Agreement. Homeless Poverty Action Coalition. URL: https://www.yolocounty.org/home/showdocument?id=55476 [accessed 2021-12-03]
46. Health home consent enrollment. New York State Department of Health. Health home consent. 2016. URL: http://www.cbcare.org/wp-content/uploads/2017/04/HHSC-CONSENT-ENROLLMENT-FOR-USE-WITH-CHILDREN-UNDER-18-DOH-5200.pdf [accessed 2021-12-03]
47. Cybersecurity and information sharing: legal challenges and solutions. Congressional Research Service. 2015. URL: https://fas.org/sgp/crs/intel/r43941.pdf [accessed 2021-12-03]

## Abbreviations

**CARES:** Coronavirus Aid, Relief, and Economic Security
**CFR:** Code of Federal Regulations
**DASH:** Data Across Sectors for Health
**FERPA:** Family Educational Rights and Privacy Act
**HIPAA:** Health Insurance Portability and Accountability Act
**HITECH:** Health Information Technology for Economic and Clinical Health Act
**HUD:** US Department of Housing and Urban Development
**GDPR:** General Data Protection Regulation
**NCHS:** National Center for Health Statistics

*Edited by R Kukafka; submitted 06.05.22; peer-reviewed by R Brown, M Doerr; comments to author 07.06.22; revised version received 17.07.22; accepted 18.07.22; published 06.09.22.*

*Please cite as:*
*Karim MA, Kum HC, Schmit CD*
*A Study of Publicly Available Resources Addressing Legal Data-Sharing Barriers: Systematic Assessment*
*J Med Internet Res 2022;24(9):e39333*
*URL: https://www.jmir.org/2022/9/e39333*
*doi:10.2196/39333*
*PMID:36066929*

Review

# Decision Support Tools in Adult Long-term Care Facilities: Scoping Review

Linda Lapp[1], MPhil; Kieren Egan[1], PhD; Lisa McCann[1], PhD; Moira Mackenzie[2], MA; Ann Wales[2], PhD; Roma Maguire[1], PhD

[1]Department of Computer and Information Sciences, University of Strathclyde, Glasgow, United Kingdom
[2]Digital Health & Care Innovation Centre, Glasgow, United Kingdom

**Corresponding Author:**
Roma Maguire, PhD
Department of Computer and Information Sciences
University of Strathclyde
26 Richmond Street
Glasgow, G1 1XH
United Kingdom
Phone: 44 141 548 3189
Email: roma.maguire@strath.ac.uk

## Abstract

**Background:** Digital innovations are yet to make real impacts in the care home sector despite the considerable potential of digital health approaches to help with continued staff shortages and to improve quality of care. To understand the current landscape of digital innovation in long-term care facilities such as nursing and care homes, it is important to find out which clinical decision support tools are currently used in long-term care facilities, what their purpose is, how they were developed, and what types of data they use.

**Objective:** The aim of this review was to analyze studies that evaluated clinical decision support tools in long-term care facilities based on the purpose and intended users of the tools, the evidence base used to develop the tools, how the tools are used and their effectiveness, and the types of data the tools use to contribute to the existing scientific evidence to inform a roadmap for digital innovation, specifically for clinical decision support tools, in long-term care facilities.

**Methods:** A review of the literature published between January 1, 2010, and July 21, 2021, was conducted, using key search terms in 3 scientific journal databases: PubMed, Cochrane Library, and the British Nursing Index. Only studies evaluating clinical decision support tools in long-term care facilities were included in the review.

**Results:** In total, 17 papers were included in the final review. The clinical decision support tools described in these papers were evaluated for medication management, pressure ulcer prevention, dementia management, falls prevention, hospitalization, malnutrition prevention, urinary tract infection, and COVID-19 infection. In general, the included studies show that decision support tools can show improvements in delivery of care and in health outcomes.

**Conclusions:** Although the studies demonstrate the potential of positive impact of clinical decision support tools, there is variability in results, in part because of the diversity of types of decision support tools, users, and contexts as well as limited validation of the tools in use and in part because of the lack of clarity in defining the whole intervention.

***(J Med Internet Res 2022;24(9):e39681)*** doi:10.2196/39681

**KEYWORDS**

decision support; care home; nursing home; digital health

## Introduction

### Background

The COVID-19 pandemic has exerted unprecedented pressure on our health and social care infrastructures. It has shown the value of rapid clinical decision-making for improving health and wellness outcomes, particularly for people in vulnerable groups (eg, older adults in care home settings). Despite the considerable potential of digital health approaches to help with continued staff shortages and improve quality of care, digital innovations are yet to make real impacts in the care home sector.

There are many barriers to implementation of clinical decision support tools in this sector, including insufficient staff and resources in care facilities, lack of carer time and knowledge [1], and limited use of electronic health records [2].

A number of reviews have been published recently regarding clinical decision support systems in long-term care facilities [2-5]. A scoping review by Abtellatif et al [2] analyzed clinical decision support systems used for pressure ulcer and malnutrition prevention, drug prescription support, and disease management, and they found 10 systems: 3 (30%) used for pressure ulcer and malnutrition prevention, 2 (20%) for medication review, 3 (30%) for daily drug prescription support, and 2 (20%) for disease management (real-time management of heart failure and management of urinary tract infection). Another systematic review by Marasinghe [5] investigated computerized clinical decision support systems used for improving medication safety. Two other systematic reviews investigated clinical decision support systems for pressure ulcer prevention and management [3,4], with the review by Araujo et al [3] assessing the effects on nurses' clinical decision-making. In addition, they investigated the factors that influence the use and successful implementation of decision support systems in clinical practice. A review by Mäki-Turja-Rostedt et al [4] explored the effectiveness of the interventions.

However, these reviews either focused only on 1 purpose of the decision support tool (eg, medication management or pressure ulcer prevention) or on a certain aspect of the support tool (eg, effectiveness only or implementation only). Hence, there is a need to better understand current evidence for the use of clinical decision support tools in long-term care facilities such as nursing and care homes. More specifically, a deeper understanding of the purpose of such tools, how these tools have been developed, and what types of data these tools use would be considerably advantageous. Therefore, the aim of this review was to analyze studies that evaluated clinical decision support tools in long-term care facilities based on (1) the purpose and intended users of the tools, (2) the evidence base used to develop the tools, (3) how the tools are used and their effectiveness, and (4) the types of data the tools use. It is anticipated that this review will contribute to existing scientific evidence to inform a roadmap for digital innovation, specifically for clinical decision support tools, in long-term care facilities.

In this review, we define *clinical decision support* as described by Greenes [6]: "Clinical decision support tools are aids for making decisions using information and communication technologies that bring relevant knowledge regarding the health and wellbeing of a patient."

### Objectives

The objective of this review was to seek answers to the following research questions (RQs):

- RQ1: What has been the purpose of clinical decision support tools? Which professionals are the intended users?
- RQ2: What evidence base was used to develop the clinical decision support tools?
- RQ3: How are clinical decision support tools used in adult long-term care facilities? What is the effectiveness of these tools?
- RQ4: What types of data do clinical decision support tools use?

To address the aforementioned questions, we undertook a scoping review [7] by reviewing recent literature (from 2010 onward), using several key search terms across 3 electronic databases.

## Methods

### Data Sources and Search Strategy

We conducted our search using a number of key search terms (("Decision Support"[tiab]) OR ("Clinical Decision-Making"[Medical Subject Headings term])) AND (("Care Home"[tiab]) OR ("Nursing Home"[tiab])) AND (2010:2021[pdat]) that were applied across 3 electronic databases: PubMed, Cochrane Library, and the British Nursing Index. Articles were included during the search if they were published between January 1, 2010, and July 31, 2021. In addition, references from included papers were screened for potential additional articles.

### Exclusion and Inclusion Criteria

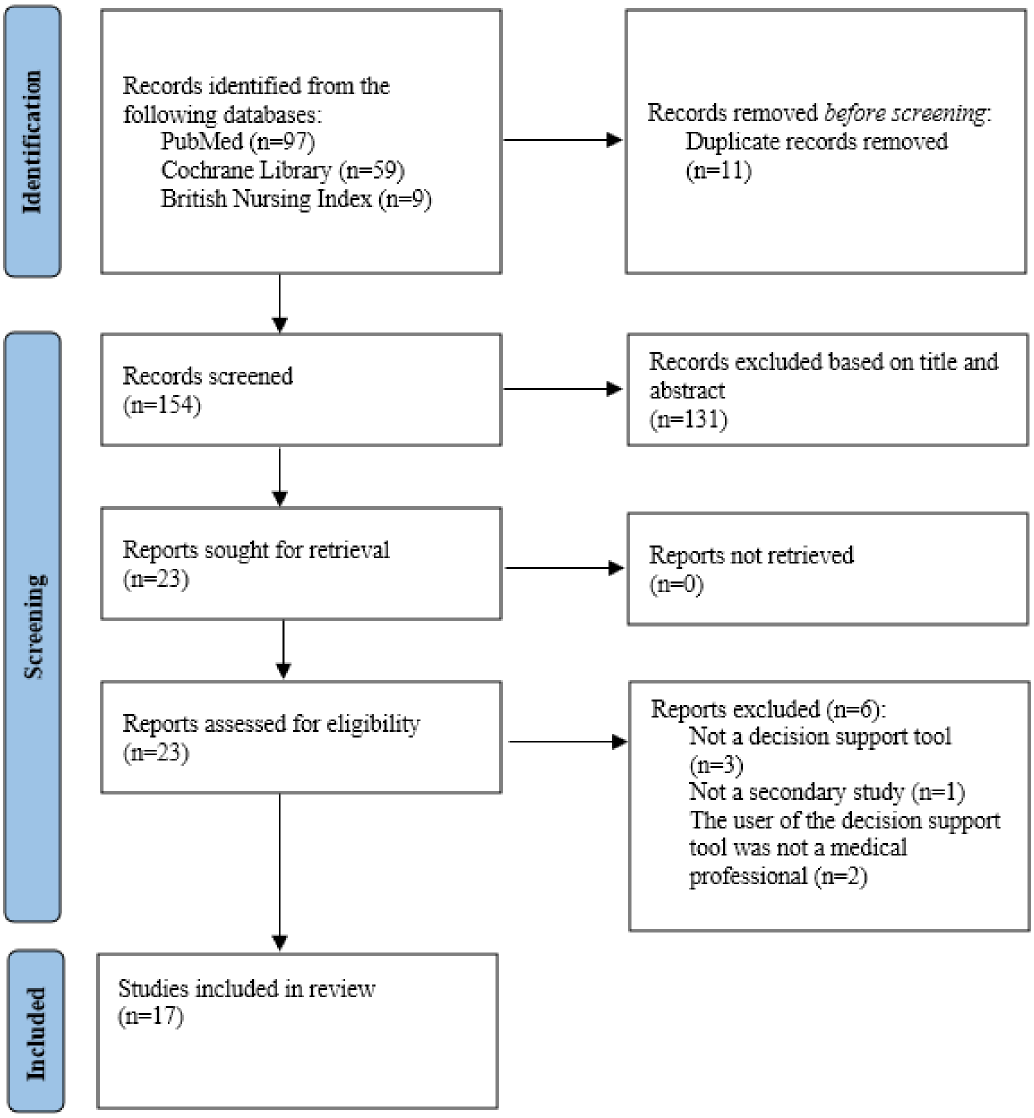
As decision support tools are an emerging area of published literature, we developed inclusion criteria (Table 1) across the parameters of setting, study design, type of decision support, user, and comparator. In terms of study design, only studies that tested or evaluated the decision support tool, as opposed to only developing the tools, were included in the review. In addition, we only included studies in English.

The exact process we used to exclude studies is further explained in the Results section and shown in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram (Figure 1).

**Table 1.** Inclusion and exclusion criteria for papers based on setting, study design, and type of decision support.

| Variable | Inclusion criteria | Exclusion criteria |
|---|---|---|
| Setting | Adult long-term care facilities (eg, adult care homes and adult nursing homes) | Nonadult long-term care facilities (eg, nonadult care homes, hospitals, and short-term care facilities) |
| Study design | The study is about testing of the decision support tool (eg, feasibility study, evaluation study, randomized controlled trial, or implementation study) | The study is developing a decision support tool without testing the tool (ie, primary study) |
| Type of decision support | The decision support tool is for patients' health conditions (both mental and physical health) | The decision support tool is for management purposes (staff planning, bed planning, etc) |
| User | Health or social care professional | Patient or family member |
| Study language | English | Other than English |

**Figure 1.** Flowchart of inclusion and exclusion of reviewed papers.



## Screening and Description of Included Studies

We screened all articles identified from the search in terms of title and abstract for potential eligibility. Where identified publications potentially met the inclusion criteria, a full text of the article was obtained for further examination. Data were collated and coded using NVivo 12 (QSR International) and Mendeley reference management software (Mendeley Ltd). Microsoft Excel was used to manage extracted data. A description of the process flow and decisions made was collated through the use of a PRISMA flow diagram [8]. Details extracted from the included publications included the following:

1. Publication details: first author of the study, year of publication, country in which the study was conducted, and scale of intervention

2. Decision support tool details: first author and year of developer publication, condition or purpose for which the tool was developed and tested, intended users of the tool, format of the tool (how it operates; eg, real-time, retrospective, or triaging system), whether it is reported to be linked to the electronic health records, whether it is reported to use a validated clinical decision support tool, and what evidence base was reported to be used to develop the tool
3. Study details: type of study (feasibility study, evaluation study, randomized controlled trial, implementation study, reflective or opinion piece, or case study), study setting, study population, study outcome, and whether there was a significant impact on the outcome from using the tool

These details of the studies have been presented in the data extraction table in Multimedia Appendix 1.

# Results

## Overview

In total, 165 papers were identified for potential inclusion; as shown in Figure 1, the search produced 97 (58.8%) papers from PubMed, 59 (35.8%) from Cochrane Library, and 9 (5.4%) from the British Nursing Index. Of these 165 papers, 11 (6.7%) were duplicates and were removed. Of the remaining 154 papers, based on title and abstract screening, 23 (14.9%) were retrieved for full-text review. Of these 23 papers, 6 (26%) were excluded for the following reasons: not studying a decision support tool (n=3, 50%), not being a study evaluating the tool (n=1, 17%), and the user of the tool not being a health or social care professional but a nursing home resident or a family member (n=2, 33%). Thus, of the 165 papers initially identified for potential inclusion, 17 (10.3%) were included in the final review [9-25].

## Setting and Study Population

All research was undertaken in high-income–country settings, including the United States (5/17, 29%), the United Kingdom (3/17, 18%), Canada (3/17, 18%), Sweden (2/17, 12%), Belgium (1/17, 6%), France (1/17, 6%), Norway (1/17, 6%), and the Netherlands (1/17, 6%).

The study population varied greatly from study to study. In studies involving care or nursing homes residents, the largest study population was 6161 residents [21], and the smallest study population was 52 residents in the study by Walker et al [24]. In studies involving health care professionals, the largest number of participants was 27 staff members in the study by Coulongeat et al [10], and the smallest number of participants was 14 registered nurses in the study by Johansson-Pajala et al [16].

## RQ1: What Has Been the Purpose of Clinical Decision Support Tools? Which Professionals Are the Intended Users?

There were 8 different conditions or purposes supported through the use of clinical decision support tools. They included medication management (5/17, 29%) [12,13,15-17], pressure ulcer prevention (4/17, 24%) [9,11,14,21], dementia management (3/17, 18%) [18-20], falls prevention (3/17, 18%) [21,23,24], hospitalization (2/17, 12%) [21,22], malnutrition prevention (1/17, 6%) [14], urinary tract infection (1/17, 6%) [25], and COVID-19 infection (1/17, 6%) [10].

In total, 65% (11/17) of the studies defined the professional user group using the decision support tool, whereas 35% (6/17) did not indicate who the intended users of the tools were [11,13,17,21,24,25]. Where stated (5/11, 45%), the most commonly specified professionals were either *care home staff* or *nursing home staff* [10,14,18,22,23]. Of these 5 studies, 4 (80%) specified that the clinical decision support tool that was evaluated was used by *nurses*; 50% (2/4) of these studies provided further detail on the types of nursing staff, including *registered nurses*, *special needs educator*, and *nurse aides* [14] or *directors, physicians, and nurses* [10]. Of the 11 studies that defined the professional user group using the decision support tool, 1 (9%) specified that the users of the decision support tool were *pharmacists* [12], and 1 (9%) stated that the tool was used by *health professionals* [16].

## RQ2: What Evidence Base Was Used to Develop the Clinical Decision Support Tools?

Overall, 88% (15/17) of the studies provided further information on how the support tools were developed. Among these 15 studies, 9 (60%) stated that the tools were developed using clinical guidelines [13,15,17,20,25-29], 11 (73%) stated that the decision support tools were developed through users' opinions [13,17,20,27-33] (eg, through the Delphi method [9,28] or other ways of stakeholder involvement [17,20,27,29,33]), 2 (13%) stated that the support tools were developed using systematic reviews of scientific evidence [34,35], and 2 (13%) used data analysis to understand the factors associated with the investigated outcome [30,32].

Of the 17 included studies, 11 (65%) evaluated the effectiveness of the decision support interventions, but no clear consensus could be arrived at on the association between the effectiveness and involvement of stakeholders or use of clinical guidelines in their development and effectiveness. Of these 11 studies, 2 (18%) evaluating tools that involved stakeholders in their development achieved significantly positive results [9,21], whereas 4 (36%) achieved mixed results [16,17,23,24], and 3 (27%) obtained nonsignificant results if the tools involved stakeholders in the development [11,20,22]. When it came to using clinical guidelines for developing decision support tools, none of the 5 studies concerned reported significant results: 4 (80%) demonstrated mixed results [16,17,23,25], and 1 (20%) obtained nonsignificant results [20].

## RQ3: How Are Clinical Decision Support Tools Used in Adult Long-term Care Facilities? What Is the Effectiveness of These Tools?

### *Medication Management*

Of the 17 included studies, 5 (29%) focused on medication management (12,13,15-17).

De Wit et al [12] evaluated a clinical decision support system that was designed for medication management. The system operates by extracting medication data of residents and 2 weeks' worth of historical laboratory data from electronic health records for all residents. Since 2008, a total of 39 clinical rules have been created in the system. If a laboratory value is deemed to be abnormal, in combination with the appropriate drug, the system generates an alert. The system then helps with dosage adjustments in accordance with various conditions, such as decreased renal function or electrolyte dysfunction [12].

The study showed that only 3 clinical rules had an efficiency of >10% *(phenytoin with hypoalbuminemia, bisphosphonates dosage regime*, and *ceftazidime with decreased renal function*). Most of the clinical rules demonstrated efficiencies of <10%, and the efficiency of 2 rules was 0% (*oral oncolytics and stop dates* and *methotrexate dosage regime*). The efficiency was calculated by dividing the number of actions for both new alerts and repeat alerts by the total number of new and repeat alerts [12]. As this was a retrospective analysis of a database, there was no control group. This means that it is difficult to evaluate whether the system had a significant effect on improved medication management.

Dorfman et al [13] investigated the potential benefits of a clinical decision support system identifying drug-gene interactions in nursing home residents who were being treated with multiple medications. They tested the system on 987 residents at 4 nursing homes. The pharmacogenetic (PGx) system uses residents' medication data and electronic health record information, together with genetic information stored in the electronic health records, to produce information regarding drug-drug interactions and other potentially dangerous drug therapy problems. On the basis of the information in the health record systems and algorithms built into the PGx system, the PGx system offers guidance to nurses and pharmacists. The study concluded that the intervention has the potential to be useful for nurses when obtaining a profile of patients' medication regarding drug-drug interactions, therapeutic duplications, and warnings for unsuitable drugs [13]. However, this is a qualitative study; therefore, no statistical significance was explored.

Johansson et al [15] evaluated the LIFe-reader, which is a PDA with a mobile medical decision support system, that was developed for safer medication management in nursing homes. The tool is used to scan the European Article Number codes on drug packages, through which the LIFe-reader generates alerts for inappropriate drugs and drug combinations: drug-drug interactions, therapeutic duplications, and warnings for drugs unsuitable for older adults. In addition to the aforementioned features, the tool includes Microsoft Word, Microsoft Excel, email, calendar, calculator, and phone. Through interviewing 22 registered nurses at various care homes, the evaluation study found that the scanning function was easy and time saving, and the LIFe-reader was useful and user friendly. However, the users requested more content and functions on the device [15].

Johansson-Pajala et al [16] studied the use of a web-based computerized decision support system that was designed for drug prescribing and medication reviews. The system is linked to electronic medical records and evaluates the quality of drug treatments based on national indicators and potential adverse drug reactions based on the residents' symptoms. The system produces warnings and explanations about inappropriate drugs, drug-drug interactions, drug use in decreased renal function, and possible adverse drug reactions. The evaluated system includes 2 widely used criteria: screening tool of older people's prescriptions, screening tool to alert to right treatment, and Beers criteria.

Kane-Gill et al [17] evaluated a clinical decision support system called TheraDoc, a clinical surveillance system containing predeveloped alerts and customizable alerts to detect potentially inappropriate prescribing, which is integrated into electronic health records. More information can be found on the TheraDoc website [36]. Alerts were created for high-risk medications, laboratory monitoring alerts, and antibiotic-stewardship–related alerts, all developed with the purpose of preventing adverse drug events. The alerts are delivered in real time [17].

The tool is reported to have been developed using medical guidelines and users' opinions; however, the study does not state exactly who the intended users of the system are. The evaluation was undertaken at 4 nursing homes, with 2127 nursing home residents as participants [17].

Of the aforementioned 5 studies focusing on medication management, only 2 (40%) were comparative studies. The study by Johansson-Pajala et al [16] showed the intervention to report significantly more adverse drug reactions and more drug-drug interactions than registered nurses. There was no significant difference between reports of inappropriate drugs and drug duplications when comparing the intervention with the actions of registered nurses. The study also investigated the nurses' views on drug management; however, the results were nonconclusive. More specifically, the registered nurses did not find that the decision support system significantly affected their drug management methods; however, many saw potential benefits of using the system [16].

The study by Kane-Gill et al [17] showed that the intervention group had significantly lower incidence of alert-specific adverse drug events than usual care. There was no statistically significant difference between the groups for all-cause hospitalizations and 30-day readmissions [17].

### *Pressure Ulcer Management and Nutrition*

Overall, 24% (4/17) of the papers focused on pressure ulcer management and nutrition [9,11,14,21].

Beeckman at al [9] evaluated an electronic clinical decision support system called PrevPlan that generates a resident-tailored protocol for pressure ulcer prevention. After data entry to the system regarding the availability of preventive materials and residents' characteristics (manual entry), the protocol included recommendations regarding skin observation, the use of support surfaces, repositioning, and heel elevation. The evaluation, which involved 464 nursing home residents and 118 health care professionals, reported that the participants had more positive than negative attitudes regarding the decision support tool, with the difference being statistically significant. It was also found that nurses with specific training regarding pressure ulcer management and higher education levels had more positive attitudes than nurses who were not experts in pressure ulcer management or were in the early years of studying to become nurses. The study found that the experimental group had significantly lower pressure ulcer incidence than the control group [9]. Further information regarding the support system can be found on the PrevPlan website [37]; however, the website was last updated in 2011 and is not available in English.

Fossum et al [14] investigated the Risk Assessment Pressure Sore (RAPS) scale for pressure ulcer risk screening and the Mini Nutritional Assessment (MNA) tool for screening nutritional status.

Both Olsho et al [21] and Davidson et al [11] evaluated *On-Time Pressure Ulcer Prevention*, which uses risk reports embedded in electronic health records to identify recent changes in risks for developing pressure ulcers. Although both studies evaluated the same decision support system, it can be assumed that these studies are not related. The system gathers information on residents' nutritional status, incontinence issues, and recent pressure ulcer history. The documentation is then used to produce 4 weekly core reports, identifying residents at high risk for pressure ulcer formation, enabling monitoring of weekly changes in risk. On-Time relies on staff communication across disciplines and documentation by certified nursing assistants. A certified nursing assistant is a person who helps patients with activities of daily living and other health care needs under the direct supervision of a registered nurse. A *change team* incorporates the reports from On-Time into clinical workflow and identifies which changes in care are required to manage the risk of developing pressure ulcers [11,21].

Olsho et al [21] carried out the evaluation, which used interrupted time-series design, at 25 nursing homes with 6161 nursing home residents as participants. The study found that the intervention components reduced pressure ulcer incidence individually and in combination [21].

Davidson et al [11] evaluated the system at 47 nursing homes; however, the study did not specify the number of participants. For their evaluation method, Davidson et al [11] used difference-in-differences design and investigated the scalability of the intervention. In spite of the large number of nursing homes involved in the study, the authors did not find whether the tool played a significant role in improving pressure ulcer prevention [11].

Of the 4 studies that investigated clinical decision support systems that aimed to help prevent pressure ulcer formation, 2 (50%) found that the incidence of pressure ulcers significantly decreased in the intervention group [9,21]; however, the remaining 2 (50%) studies found no statistically significant differences in terms of pressure ulcer incidence [11,14]. Only the study by Fossum et al [14] used a validated tool (RAPS) for pressure ulcer management. In their study, Fossum et al [14] did not find significant differences in nutritional status of residents between the intervention and control groups when using the MNA.

### *Dementia Management*

In total, 18% (3/17) of the studies focused on evaluating decision support tools developed for dementia management. Keenan et al [18] and Moniz-Cook et al [20] evaluated 2 different decision support systems that aimed to help care home staff support residents with commonly occurring challenging behaviors. The system included assessment tools that collected relevant information regarding the residents and then applied logic-based algorithms that generated biopsychosocial action plans that the staff could implement [18,20].

The study by Keenan et al [18] was a qualitative study looking at the contextual and organizational mechanisms of, as well as barriers and facilitators for, the intervention. Four mechanisms of implementation of the intervention were identified: (1) access to, and use of, care homes; (2) resources in terms of IT for e-learning activity; (3) demonstrating capacity to apply action care planning in care practice; and (4) receptivity of care home staff to e-learning and the individually tailored action care planning that followed [18].

The study by Moniz-Cook et al [20] was a cluster randomized trial undertaken at 63 care homes. In total, 658 nursing home residents and 436 care home staff members took part in the study. It was found that there was no statistically significant difference in the number of incidents of challenging behavior between the intervention and control groups. The intervention did not significantly affect the experience of staff members with regard to the prescription of psychotropic medication. The quality of life of residents was not measured because of the large amount of missing data [20].

Kovach et al [19] evaluated a decision support tool called the Serial Trial Intervention, which aimed to help with assessment and treatment of pain and other physical problems of residents with advanced dementia who are unable to report symptoms clearly or consistently. The tool is a 9-step assessment and treatment process, previously evaluated as a 5-step tool. If an assessment is negative, or if interventions fail to decrease symptoms, the nurse moves to the next step. The study, undertaken at 12 nursing homes with 125 nursing home residents, compared the effectiveness of the 2 versions of the protocol. It found that the residents being treated using the 9-step intervention received more assessment-driven treatment and evaluation-driven follow-up. It was also found that these residents had less static and dismissive care than those treated using the 5-step intervention [19].

### *Falls Prevention*

Of the 17 included studies, 2 (12%) focused on evaluations of interventions that aimed to support the management or prevention of falls. Tzeng et al [23] carried out a quality improvement project to evaluate the impact of the Fall Tailoring Interventions for Patient Safety program on preventing falls and fall-related injuries among residents. This program was developed to help staff modify falls prevention interventions based on daily assessments, and it can be used as a personalized falls prevention plan displayed on screens placed at residents' bedsides. The program was developed to help nursing staff identify evidence-based interventions for each area of risk. In the paper, Tzeng et al [23] report that after implementing the intervention the reduction in the average monthly fall rate was clinically significant: the average monthly fall rate reduced from 10.07 falls to 7.95 falls. However, no statistical significance was reported [23].

Walker et al [24] evaluated the Guide to Action Care Home falls prevention program, which consists of posters and paper-based decision support tools in the form of a checklist that helps to identify risk factors associated with falls and suggests actions to reverse or modify these falls risk factors. Walker et al [24] found that the fall rates were lower, and there were nearly twice as many general practitioner visits at control homes than at intervention homes over 6 months of follow-up. That being said, no statistical significance was reported because of the small number of falls [24].

### Other Purposes: Hospitalization, Urinary Tract Infection, and COVID-19 Infection

Pasay et al [25] evaluated a decision support tool based on the principles of building a culture of safe, effective, and sustainable antimicrobial use for urinary tract infection. The intervention consisted of 4 parts: education of physicians, nursing staff, families, and caregivers; posters with myths and facts regarding the diagnosis and treatment of urinary tract infection; a pamphlet for family and caregivers; and a clinical tool to help with behavioral changes in residents (drugs, eyes and ears, low oxygen states, infection, retention of urine or stool, ictal, underhydration or undernutrition, metabolic, and subdural [DELIRIUMS] tool). The evaluated decision support tool is a checklist that guides staff to identify urinary tract infections based on clinical symptoms, to collect a urine culture only when indicated, and to review antimicrobial therapy if prescribed. The checklist also acted as an interprofessional communication tool [25].

Pasay et al [25] found that there was a statistically significant reduction in urine testing in the intervention group compared with the control group. There was also a statistically significant reduction in the rates of antimicrobial prescribing in the intervention group compared with the control group. There were no differences in admissions to acute care or the emergency department between the 2 groups [25].

Tena-Nelson et al [22] evaluated a program called Interventions to Reduce Acute Care Transfers New York, which consisted of six parts: (1) the Situation, Background, Assessment, Recommendation (SBAR) tool, which was designed to help with communication among medical professionals; (2) the Early Warning Tool (Stop and Watch), which was designed to help in recognizing significant change in residents early; (3) a hospital transfer review tool to guide retrospective review of hospitalizations; (4) a standardized patient transfer form and a checklist; (5) care paths to guide treatment options for common conditions; and (6) advance care planning tools to guide decision-making and communication about end-of-life care.

SBAR is a structured communication format that enables health information to be transferred between individuals and institutions. It aims to convey critical information understandably, clearly, and succinctly [38].

Stop and Watch is a tool to help spot the signs warning that a person's condition is deteriorating. The poster helps staff to recognize signs and take steps to reduce a person's risk of morbidity, further disability, organ failure, and mortality [39].

According to the study by Tena-Nelson et al [22], there were no statistically significant changes in hospitalization rates between before and after the intervention. No statistically significant factors were found to be associated with the changes in hospitalization rate. The authors stipulated that the program's effectiveness could be improved by including participant recommendations on planning, staff and stakeholder engagement, implementation, training, and sustainability [22].

Coulongeat et al [10] investigated a local support platform that aimed to help nursing homes manage their cases of COVID-19 infection. Although the other decision support tools included in the review consisted mostly of 2 actors (a human and a computer [or paper in some cases]), the COVID-19–infection management tool used multiple human actors as the decision support. The reason for this might be that because COVID-19, at the time of tool development, was a very new disease, there was very little evidence available to support the development of a computer program that could help manage the condition. The decision support aspect consisted of a multidisciplinary team, a specialist phone hotline, and mobile geriatric medicine teams, all reachable through information and communication technologies. The intervention helped to satisfactorily address some issues that were revealed by the COVID-19 pandemic. These issues were as follows: limiting the feeling of isolation, getting the health professionals' questions answered, providing solutions to individual problems, and reassurance of the nursing home staff regarding the optimal treatments for residents. The intervention was less effective in improving the quality of life for residents or staff at nursing homes with a COVID-19 cluster [10]; however, why this was the case is not explained in the paper.

## RQ4: What Types of Data Do Clinical Decision Support Tools Use?

### Integration With Electronic Health Records

Of the 17 included studies, only 4 (24%) described decision support tools that were integrated with an electronic health record, all of which were developed for medication management [12,13,16,17]. However, it is not known what these electronic health records consist of, whether these are stand-alone systems for the care facilities, or whether these records are linked with general practices. On the basis of the information presented, it is known that in 12% (2/17) of the included papers, the decision support tools were not linked to any electronic health records because they were paper-based posters [23,24]. For the remaining 65% (11/17) of the studies, it is unknown whether the decision support tools were stand-alone systems or linked to electronic health records.

### Data Used by Clinical Decision Support Tools

Overall, based on their original studies, it was unclear what kind of information the decision support tools required. In total, 29% (5/17) of the studies were clear about the data being collected to aid decision-making. Of these 5 studies, 2 (40%) described decision support tools used for dementia management (Serial Trial Intervention [19] and DemCare [20]), 3 (60%) described decision support tools used for pressure ulcer prevention and management (On-Time [11,21] and RAPS [14]) as well as malnutrition prevention (MNA [14]), and 1 (20%) described a decision support tool used for urinary tract infection prevention and management (multimodal antimicrobial stewardship intervention [25]).

## Discussion

### Overview

In this review we set out to collate current knowledge within the academic literature focused on decision support tools in long-term care settings. Decision support tools are an emerging area of research and practice spanning a range of different conditions, health, and social care professions. However, many studies to date have focused on small-scale, localized efforts; only 17 studies conducted in only 8 high-income–country settings were found to have evaluated decision support tools developed for use in long-term care facilities. Furthermore, although a small number (3/17, 18%) of the identified studies present favorable outcomes, this was not universally true, and there is often a reliance on early evidence such as short-term evaluation studies and analysis of qualitative data. Thus, although this area of research holds significant potential, our findings suggest that review of the published literature is timely to inform future innovation.

We are now moving to a data-driven health and social care model; therefore, the concept of siloed data needs to be a thing of the past, and available data must be used for the benefit of residents of nursing and care homes and to provide added value.

### Principal Findings

In terms of setting and study sample, the majority (13/17, 76%) of the included studies were local, with only 24% (4/17) being carried out on a national scale. That being said, the majority (15/17, 88%) of the studies were multicenter studies, with only 12% (2/17) [23,40] being single-center studies. In terms of the clinical populations who were using the decision support systems at adult long-term care facilities, the most common were staff members, nurses, pharmacists, and health professionals in general. However, the studies in general did not explicitly specify who the intended users were. In addition, 35% (6/17) of the studies did not indicate who the intended users of the tools were.

It should be noted that it is uncommon for long-term care facilities, such as care homes and nursing homes, for example, in the United Kingdom, to have a physician or pharmacist present. Instead, care facilities have partnerships with local general practitioner practices. Hence, it is unlikely that staff within care or nursing homes will be making decisions about medications [41].

Although there is potential for clinical decision support tools to streamline care services by making them more efficient, the systems have been developed to address issues that often fall under the clinical responsibility of nursing staff; therefore, as expected, most users fall within this profession. However, lack of detail regarding users does limit insights into the implementation of the system and therefore may be deemed to inhibit the transferability and scaling up of these systems to other sites and domains [42].

Considering the average age of long-term care residents, the conditions of focus for the clinical decision support tools are not surprising. There were 8 different conditions or purposes supported using clinical decision support tools: medication management, pressure ulcer prevention, dementia management, falls prevention, hospitalization, malnutrition prevention, urinary tract infection, and COVID-19 infection. It is noteworthy that all tools seem to focus on domains of physical health; none focused on mental health, despite it being widely recognized that rates of anxiety and depression are high in this population [43].

Looking at studies demonstrating evidence of whether there was a significant improvement through using the clinical decision support tools, 71% (12/17) of the studies carried out comparative analyses. Of these 12 studies, only 3 (25%) reported clear significance in the results, showing that the evaluated decision support tools made a difference in either preventing negative outcomes or improving care in general [9,19,21].

Of these 3 studies, 2 (67%) evaluated clinical decision support tools developed for pressure ulcer management [9,21]. Olsho et al [21] and Davidson et al [11] evaluated the same tool for pressure ulcer management; however, Davidson et al [11] reported no statistically significant change in the incidence of pressure ulcers when the intervention was used. According to a systematic review by Mäki-Turja-Rostedt et al [4], there are many ways to prevent pressure ulcer formation in residents of long-term care facilities; however, there is a lack of systematic evidence of the most effective way to do this. In their systematic review, Araujo et al [3] agree and add that clinical effects, such as outcomes in the incidence and prevalence of pressure ulcers, remain limited, and most investigated studies found clinically but nonstatistically significant results in decreasing pressure ulcer incidence. The results from this review are in concordance with the comments made in these systematic reviews [3,4].

Of the 17 studies, 5 (29%) found mixed results, meaning that the intervention improved some outcomes but reported no statistically significant difference in other outcomes [16,17,23-25]. However, it is important to note that 40% (2/5) of these studies mentioned “clinical significance” but did not define how this clinical significance was measured and did not report on statistical significance [23,24]. In total, 24% (4/17) of the studies included in the review showed no statistically significant difference when using the intervention [11,14,20,22]. This lack of definitive evidence underpinning digital health solutions is widely recognized; often, tools are implemented within large organizations with very little evidence underpinning them [2,5]. This lack of evidence underscores the need for robust evaluation of solutions used to identify benefits of, and value of investment in, decision support tools [44].

It is important to note that, as explained previously and shown in the data extraction table (Multimedia Appendix 1), there is a noticeable degree of heterogeneity among the decision support tools described in the included studies in terms of their purposes and intended users. However, even among decision support tools with similar purposes, the study designs and measured outcomes in the included evaluation studies varied substantially. Hence, these studies should be compared with one another with extreme caution. Because of this variability, further studies, using standardized methods to evaluate the decision support tools included in this review, are needed.

### Evidence Base of the Development of Clinical Decision Support Tools

An important question that this review aimed to answer concerns the evidence base underpinning the development of decision support tools. Our findings demonstrate that there is a lack of coherent information about what evidence base was used to develop the clinical decision support tools. Most tools were developed based on current guidelines and stakeholders' opinions. Some studies used systematic reviews of scientific evidence or data analysis to understand the factors associated with the outcome of the tool. In 12% (2/17) of the studies, there is no information regarding the evidence base of the development of the tools. The lack of transparency in underpinning evidence of development of the tools can affect users' trust in the tools, ultimately affecting their wider uptake [42].

### Adoption and Implementation of Clinical Decision Support Tools

For successful implementation of clinical decision support tools, educational training and culture change are required to sustain their clinical use. Clinical decision support tools, especially those helping to manage medications, need to be regularly updated with regard to changing guidelines and newly available drugs. Ease of updating is an important factor for being considered a successful decision support tool [2,5,38].

A key consideration in terms of the mechanisms and actions of the clinical decision support tools is the data that they use. Overall, in the included studies, all tools, apart from the paper-based decision support tools (n=2), are computer programs. However, most (11/17, 65%) of the studies were not explicit about whether the computer programs were linked to electronic health records, and if they were linked (4/17, 24%), then what kind of electronic health records they were linked to.

Of the 17 included studies, only 3 (18%) described decision support tools that used validated clinical tools. These validated clinical tools are RAPS and MNA used in the evaluation study by Fossum et al [14]; screening tool of older people's prescriptions, screening tool to alert to right treatment, and Beers criteria used in the evaluation study by Johansson-Pajala et al [16]; and SBAR and Stop and Watch used in the evaluation study by Tena-Nelson et al [22]. When considering the studies (4/17, 24%) that evaluated clinical decision support tools helping to prevent pressure ulcer formation, the study by Fossum et al [14] was the only one to use validated clinical tools.

Considering how long electronic health records have existed and how long researchers have been working on predictive modeling in health, it is surprising that more widely validated decision support tools were not included in the evaluation studies.

It can be assumed that the evaluated tools were developed keeping in mind that electronic health records are not widely used in long-term care facilities such as care and nursing homes. Most of the tools presented in this review, except for medication management tools, do not necessarily require data that are normally stored in electronic health records (eg, temporary symptoms such as meal intake and urinary frequency). Hence, these tools require manual data input from staff, which requires developing suitable infrastructure for tool use (eg, the availability of computers or tablets) at the facilities and training of, and time from, staff. It has been found that these can affect methodological challenges when validating clinical tools, such as gaining acceptance from stakeholders who have limited time in addition to having to deal with work pressure, such as registered nurses, care home managers, and general practitioners [45].

If long-term care facilities would adopt electronic health records specific to these facilities, the burden of having to manually input data for each different decision support tool would be removed, and the tools can be integrated into the practice automatically. This has been successfully implemented in various hospital settings. The systematic review by Varghese et al [46] provides examples of such systems successfully implemented in hospitals.

It is important to understand, however, that hospitals are data-rich environments with often automated data collection (eg, intensive care units), and the purpose of the constant data collection is due to having to always monitor patients. In addition, it is worth noting that having a multiplicity of duplicative patient records is not the aim; rather, the goal is to create a shared patient record using approved data sources that all services can access. Long-term care facilities such as care and nursing homes are essentially the residents' homes, meaning that to be able to provide a homely and comfortable setting, a balanced approach regarding data collection needs to be achieved. On the one hand, having electronic health records as in a hospital setting would open up opportunities for more decision support tools that could potentially reduce the workload of staff and help to improve care quality. On the other hand, to offer as comfortable a living environment as possible to residents, data collection should be limited to only those occasions when residents are ill or are at risk of an illness.

To strike a balance, minimum data sets could be the answer. A review of uptake of minimum data sets by Musa et al [47] evaluated different contexts, mechanisms, and outcomes to describe why minimum data sets were used in care homes, including system-level, care home–level, and individual-level barriers and facilitators. Some of the barriers mentioned include frequent staff turnover, training issues, and lack of computer skills. Facilitators include clinical staff presence, inclusive and understanding care home culture, and clarity of roles in data collection. These are very similar to barriers and facilitators for the implementation of clinical decision support tools that were identified in the studies in this review. The factors associated with successful system implementation in clinical practice were as follows: involving the administrator or head of nursing in the process; engaging the leadership in the project; presence of an internal champion; and participation of an interdisciplinary team, facilitators, and quality improvement team. In addition, it was deemed necessary to consider clinical workflow and training needs. It was also recommended to have a longer evaluation period to assess the effect of clinical decision support systems [3].

## Conclusions

Overall, the studies demonstrate that decision support tools can show improvements in delivery of care and in health outcomes, specifically in relation to medication management, falls prevention, management of dementia, pressure ulcer prevention and management, and nutritional assessment and management. However, there is variability in results, in part because of the diversity of types of decision support tools, users, and contexts as well as limited validation of the tools in use and in part because of the lack of clarity in defining the whole intervention. An important aspect that the studies seem to highlight is that decision-making to support care home residents is not just about providing technology within care homes; it also requires an effective multiagency approach with interaction with the wider multidisciplinary team outside the care home and supportive organization and culture to embed the use of the decision support tools.

## Acknowledgments

This study was part of the Care Homes Assessment Tool Stage 2: "Proof of Concept" project funded by the Scottish Government and Digital Health & Care Innovation Centre.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Data extraction table.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v24i9e39681_app1.xlsx ]

## References

1. White EM, Aiken LH, McHugh MD. Registered nurse burnout, job dissatisfaction, and missed care in nursing homes. J Am Geriatr Soc 2019 Oct;67(10):2065-2071 [FREE Full text] [doi: 10.1111/jgs.16051] [Medline: 31334567]
2. Abdellatif A, Bouaud J, Nghiem D, Lafuente-Lafuente C, Belmin J, Seroussi B. Clinical decision support systems in nursing homes: a scoping review. Stud Health Technol Inform 2020 Jun 16;270:542-546. [doi: 10.3233/SHTI200219] [Medline: 32570442]
3. Araujo SM, Sousa P, Dutra I. Clinical decision support systems for pressure ulcer management: systematic review. JMIR Med Inform 2020 Oct 16;8(10):e21621 [FREE Full text] [doi: 10.2196/21621] [Medline: 33064099]
4. Mäki-Turja-Rostedt S, Stolt M, Leino-Kilpi H, Haavisto E. Preventive interventions for pressure ulcers in long-term older people care facilities: a systematic review. J Clin Nurs 2019 Jul 25;28(13-14):2420-2442. [doi: 10.1111/jocn.14767] [Medline: 30589987]
5. Marasinghe KM. Computerised clinical decision support systems to improve medication safety in long-term care homes: a systematic review. BMJ Open 2015 May 12;5(5):e006539 [FREE Full text] [doi: 10.1136/bmjopen-2014-006539] [Medline: 25967986]
6. Greenes R. Definition, scope and challenges. In: Clinical Decision Support The Road Ahead. Cambridge, Massachusetts, United States: Academic Press; 2007.
7. Sucharew H, Macaluso M. Progress notes: methods for research evidence synthesis: the scoping review approach. J Hosp Med 2019 Jul 01;14(7):416-418. [doi: 10.12788/jhm.3248] [Medline: 31251164]
8. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. 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]
9. Beeckman D, Clays E, Van Hecke A, Vanderwee K, Schoonhoven L, Verhaeghe S. A multi-faceted tailored strategy to implement an electronic clinical decision support system for pressure ulcer prevention in nursing homes: a two-armed randomized controlled trial. Int J Nurs Stud 2013 Apr;50(4):475-486. [doi: 10.1016/j.ijnurstu.2012.09.007] [Medline: 23036149]
10. Coulongeat M, Aïdoud A, Poupin P, Dubnitskiy-Robin S, Fougère B. A local COVID-19 support platform for nursing homes: feedback and perspectives. J Am Med Dir Assoc 2021 Aug;22(8):1599-1601 [FREE Full text] [doi: 10.1016/j.jamda.2021.06.001] [Medline: 34197795]
11. Davidson C, Loganathan S, Bishop L, Imhof L, Bergofsky L, Spector W, et al. Scalability of an IT intervention to prevent pressure ulcers in nursing homes. J Am Med Dir Assoc 2019 Jul;20(7):816-21.e2. [doi: 10.1016/j.jamda.2019.02.008] [Medline: 30954418]
12. de Wit HA, Mestres Gonzalvo C, Cardenas J, Derijks HJ, Janknegt R, van der Kuy PM, et al. Evaluation of clinical rules in a standalone pharmacy based clinical decision support system for hospitalized and nursing home patients. Int J Med Inform 2015 Jun;84(6):396-405. [doi: 10.1016/j.ijmedinf.2015.02.004] [Medline: 25746461]

13. Dorfman R, London Z, Metias M, Kabakchiev B, Mukerjee G, Moser A. Individualized medication management in Ontario long-term care clinical impact on management of depression, pain, and dementia. J Am Med Dir Assoc 2020 Jun;21(6):823-9.e5. [doi: 10.1016/j.jamda.2020.04.009] [Medline: 32536434]
14. Fossum M, Alexander GL, Ehnfors M, Ehrenberg A. Effects of a computerized decision support system on pressure ulcers and malnutrition in nursing homes for the elderly. Int J Med Inform 2011 Sep;80(9):607-617. [doi: 10.1016/j.ijmedinf.2011.06.009] [Medline: 21783409]
15. Johansson PE, Petersson GI, Nilsson GC. Personal digital assistant with a barcode reader-a medical decision support system for nurses in home care. Int J Med Inform 2010 Apr;79(4):232-242. [doi: 10.1016/j.ijmedinf.2010.01.004] [Medline: 20138577]
16. Johansson-Pajala R, Martin L, Jorsäter Blomgren K. Registered nurses' use of computerised decision support in medication reviews. Int J Health Care Qual Assur 2018 Jul 09;31(6):531-544. [doi: 10.1108/IJHCQA-01-2017-0009] [Medline: 29954263]
17. Kane-Gill SL, Wong A, Culley CM, Perera S, Reynolds MD, Handler SM, et al. Transforming the medication regimen review process using telemedicine to prevent adverse events. J Am Geriatr Soc 2021 Feb;69(2):530-538. [doi: 10.1111/jgs.16946] [Medline: 33233016]
18. Keenan J, Poland F, Manthorpe J, Hart C, Moniz-Cook E. Implementing e-learning and e-tools for care home staff supporting residents with dementia and challenging behaviour: a process evaluation of the ResCare study using normalisation process theory. Dementia (London) 2020 Jul;19(5):1604-1620 [FREE Full text] [doi: 10.1177/1471301218803195] [Medline: 30269534]
19. Kovach CR, Simpson MR, Joosse L, Logan BR, Noonan PE, Reynolds SA, et al. Comparison of the effectiveness of two protocols for treating nursing home residents with advanced dementia. Res Gerontol Nurs 2012 Oct;5(4):251-263 [FREE Full text] [doi: 10.3928/19404921-20120906-01] [Medline: 22998656]
20. Moniz-Cook E, Hart C, Woods B, Whitaker C, James I, Russell I, et al. Challenge Demcare: management of challenging behaviour in dementia at home and in care homes – development, evaluation and implementation of an online individualised intervention for care homes; and a cohort study of specialist community mental health care for families. Programme Grants Appl Res 2017 Aug;5(15):1-290. [doi: 10.3310/pgfar05150] [Medline: 28783270]
21. Olsho LE, Spector WD, Williams CS, Rhodes W, Fink RV, Limcangco R, et al. Evaluation of AHRQ's on-time pressure ulcer prevention program: a facilitator-assisted clinical decision support intervention for nursing homes. Med Care 2014 Mar;52(3):258-266. [doi: 10.1097/MLR.0000000000000080] [Medline: 24374408]
22. Tena-Nelson R, Santos K, Weingast E, Amrhein S, Ouslander J, Boockvar K. Reducing potentially preventable hospital transfers: results from a thirty nursing home collaborative. J Am Med Dir Assoc 2012 Sep;13(7):651-656. [doi: 10.1016/j.jamda.2012.06.011] [Medline: 22835484]
23. Tzeng H, Jansen LS, Okpalauwaekwe U, Khasnabish S, Andreas B, Dykes PC. Adopting the fall tailoring interventions for patient safety (TIPS) program to engage older adults in fall prevention in a nursing home. J Nurs Care Qual 2021;36(4):327-332. [doi: 10.1097/NCQ.0000000000000547] [Medline: 33534349]
24. Walker GM, Armstrong S, Gordon AL, Gladman J, Robertson K, Ward M, et al. The Falls In Care Home study: a feasibility randomized controlled trial of the use of a risk assessment and decision support tool to prevent falls in care homes. Clin Rehabil 2016 Oct;30(10):972-983 [FREE Full text] [doi: 10.1177/0269215515604672] [Medline: 26385358]
25. Pasay DK, Guirguis MS, Shkrobot RC, Slobodan JP, Wagg AS, Sadowski CA, et al. Antimicrobial stewardship in rural nursing homes: impact of interprofessional education and clinical decision tool implementation on urinary tract infection treatment in a cluster randomized trial. Infect Control Hosp Epidemiol 2019 Apr;40(4):432-437. [doi: 10.1017/ice.2019.9] [Medline: 30782227]
26. Eppenga WL, Derijks HJ, Conemans JM, Hermens WA, Wensing M, De Smet PA. Comparison of a basic and an advanced pharmacotherapy-related clinical decision support system in a hospital care setting in the Netherlands. J Am Med Inform Assoc 2012 Jan 01;19(1):66-71 [FREE Full text] [doi: 10.1136/amiajnl-2011-000360] [Medline: 21890873]
27. Gallagher PF, O'Connor MN, O'Mahony D. Prevention of potentially inappropriate prescribing for elderly patients: a randomized controlled trial using STOPP/START criteria. Clin Pharmacol Ther 2011 Jun 20;89(6):845-854. [doi: 10.1038/clpt.2011.44] [Medline: 21508941]
28. Hamilton H, Gallagher P, Ryan C, Byrne S, O'Mahony D. Potentially inappropriate medications defined by STOPP criteria and the risk of adverse drug events in older hospitalized patients. Arch Intern Med 2011 Jun 13;171(11):1013-1019. [doi: 10.1001/archinternmed.2011.215] [Medline: 21670370]
29. Dykes PC, Duckworth M, Cunningham S, Dubois S, Driscoll M, Feliciano Z, et al. Pilot testing fall TIPS (tailoring interventions for patient safety): a patient-centered fall prevention toolkit. Jt Comm J Qual Patient Saf 2017 Aug;43(8):403-413. [doi: 10.1016/j.jcjq.2017.05.002] [Medline: 28738986]
30. Beeckman D, Defloor T, Demarré L, Van Hecke A, Vanderwee K. Pressure ulcers: development and psychometric evaluation of the attitude towards pressure ulcer prevention instrument (APuP). Int J Nurs Stud 2010 Nov;47(11):1432-1441. [doi: 10.1016/j.ijnurstu.2010.04.004] [Medline: 20466370]

31. Sharkey S, Hudak S, Horn S, Spector W. Leveraging certified nursing assistant documentation and knowledge to improve clinical decision making: the on-time quality improvement program to prevent pressure ulcers. Adv Skin Wound Care 2011 Apr;24(4):182-8; quiz 188. [doi: 10.1097/01.ASW.0000396244.65415.7f] [Medline: 21407045]
32. Ouslander JG, Bonner A, Herndon L, Shutes J. The Interventions to Reduce Acute Care Transfers (INTERACT) quality improvement program: an overview for medical directors and primary care clinicians in long term care. J Am Med Dir Assoc 2014 Mar;15(3):162-170 [FREE Full text] [doi: 10.1016/j.jamda.2013.12.005] [Medline: 24513226]
33. Robertson K, Logan P, Ward M, Pollard J, Gordon A, Williams W, et al. Thinking falls-taking action: a falls prevention tool for care homes. Br J Community Nurs 2012 May;17(5):206-209. [doi: 10.12968/bjcn.2012.17.5.206] [Medline: 22584396]
34. Lindgren M, Unosson M, Krantz A, Ek A. A risk assessment scale for the prediction of pressure sore development: reliability and validity. J Adv Nurs 2002 Apr;38(2):190-199. [doi: 10.1046/j.1365-2648.2002.02163.x] [Medline: 11940132]
35. Guigoz Y, Vellas B. The Mini Nutritional Assessment (MNA) for grading the nutritional state of elderly patients: presentation of the MNA, history and validation. Nestle Nutr Workshop Ser Clin Perform Programme 1999;1:3-11; discussion 11. [doi: 10.1159/000062967] [Medline: 11490593]
36. Theradoc homepage. Theradoc. URL: https://www.theradoc.com/ [accessed 2022-03-30]
37. Beeckman D, Vanderwee K, Schoonhoven L, Defloor T. PrevPlan homepage. PrevPlan. URL: http://www.decubitus.be/PrevPlan/ [accessed 2022-03-30]
38. Shahid S, Thomas S. Situation, background, assessment, recommendation (SBAR) communication tool for handoff in health care – a narrative review. Saf Health 2018 Jul 28;4(1). [doi: 10.1186/s40886-018-0073-1]
39. Lee C, Gruss V, Stuercke M, Ryan C. Relation of stop and watch tool use in a skilled nursing facility to 30-day hospital transfer rates. Jour Nursing Home Res 2016:104-109. [doi: 10.14283/jnhrs.2016.16]
40. Houthooft R, Ruyssinck J, van der Herten J, Stijven S, Couckuyt I, Gadeyne B, et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. Artif Intell Med 2015 Mar;63(3):191-207. [doi: 10.1016/j.artmed.2014.12.009] [Medline: 25579436]
41. National Institute for Health and Care Excellence (NICE). In: The Grants Register 2018. London: Palgrave Macmillan; 2018.
42. Alexander GL. Issues of trust and ethics in computerized clinical decision support systems. Nurs Adm Q 2006;30(1):21-29. [doi: 10.1097/00006216-200601000-00005] [Medline: 16449881]
43. O Neill M, Ryan A, Slater P, Ferry F, Bunting B. Mental health, quality of life and medication use among care home residents and community dwelling older people. Int J Res Nurs 2019 Jan 01;10(1):10-23. [doi: 10.3844/ijrnsp.2019.10.23]
44. How do digital health standards assess evidence? ORCHA. 2021 Jul 14. URL: https://orchahealth.com/how-do-digital-health-standards-assess-evidence/ [accessed 2022-04-07]
45. Brennan CW, Daly BJ. Methodological challenges of validating a clinical decision-making tool in the practice environment. West J Nurs Res 2015 Apr 18;37(4):536-545. [doi: 10.1177/0193945914539738] [Medline: 24948588]
46. Varghese J, Kleine M, Gessner S, Sandmann S, Dugas M. Effects of computerized decision support system implementations on patient outcomes in inpatient care: a systematic review. J Am Med Inform Assoc 2018 May 01;25(5):593-602 [FREE Full text] [doi: 10.1093/jamia/ocx100] [Medline: 29036406]
47. Musa MK, Akdur G, Brand S, Killett A, Spilsbury K, Peryer G, et al. The uptake and use of a minimum data set (MDS) for older people living and dying in care homes: a realist review. BMC Geriatr 2022 Jan 07;22(1):33 [FREE Full text] [doi: 10.1186/s12877-021-02705-w] [Medline: 34996391]

## Abbreviations

**DELIRIUMS:** drugs, eyes and ears, low oxygen states, infection, retention of urine or stool, ictal, underhydration or undernutrition, metabolic, and subdural
**MNA:** Mini Nutritional Assessment
**PGx:** pharmacogenetic
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RAPS:** Risk Assessment Pressure Sore
**RQ:** research question
**SBAR:** Situation, Background, Assessment, Recommendation

*Edited by R Kukafka; submitted 18.05.22; peer-reviewed by RM Johansson-Pajala, R Mbiine; comments to author 23.06.22; revised version received 14.07.22; accepted 11.08.22; published 06.09.22.*

*Please cite as:*
*Lapp L, Egan K, McCann L, Mackenzie M, Wales A, Maguire R*
*Decision Support Tools in Adult Long-term Care Facilities: Scoping Review*
*J Med Internet Res 2022;24(9):e39681*
*URL: https://www.jmir.org/2022/9/e39681*
*doi:10.2196/39681*
*PMID:36066928*

Review

# Telemedicine for Preventing and Treating Pressure Injury After Spinal Cord Injury: Systematic Review and Meta-analysis

Guilian Chen[1], MSN; Tong Wang[2], BSN; Lirong Zhong[1], BSN; Xinghui He[1], BSN; Chunxia Huang[1], BSN; Yingmin Wang[1*], MSN; Kun Li[2*], PhD

[1]Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou, China
[2]School of Nursing, Sun Yat-Sen University, Guangzhou, China
[*]these authors contributed equally

**Corresponding Author:**
Kun Li, PhD
School of Nursing
Sun Yat-Sen University
74, Zhongshan 2nd Road
Yuexiu District
Guangzhou
China
Phone: 86 020 87331922
Fax: 86 020 87333043
Email: likun22@mail.sysu.edu.cn

## Abstract

**Background:** Pressure injury is a common complication after a spinal cord injury. Long-term multidisciplinary follow-up is difficult after such patients have been discharged. Telemedicine promises to provide convenient and effective support for the prevention and treatment of pressure injury, but previous attempts to demonstrate that have produced inconsistent results.

**Objective:** The aim of this study is to evaluate the effectiveness of telemedicine in preventing and treating pressure injury among community-dwelling patients with spinal cord injury, and determine which telemedicine form is more effective.

**Methods:** This systematic review was performed according to the PRISMA-NMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Network Meta-Analysis) standards. Ten databases were searched to identify randomized controlled trials and quasi-experimental studies related to the effectiveness of telemedicine intervention in patients with spinal cord injury. Two researchers worked independently and blindly selected studies, extracted data, and assessed the risk of bias. The results were described as relative risk (RR) and weighted mean difference and 95% CI.

**Results:** The 35 studies comprised 25 randomized controlled trials and 10 quasi-experimental studies involving 3131 patients. The results showed that telemedicine can significantly ($P<.05$) reduce the incidence of pressure injury (RR 0.24, 95% CI 0.14-0.41; $P<.05$; $I^2=0\%$), promote faster healing (RR 0.73, 95% CI 0.62-0.85; $P<.05$; $I^2=0\%$), and yield lower scores on the pressure ulcer scale of healing (weighted mean difference=–1.98, 95% CI –3.51 to –0.46; $P<.05$; $I^2=0\%$). Cumulative ranking estimates showed that combining telemedicine with conventional intervention (93.5%) was the most effective approach.

**Conclusions:** Telemedicine is a feasible way to prevent pressure injury among patients with spinal cord injuries. It can decrease the incidence and severity of pressure injury and accelerate patients’ healing without imposing economic burden. It is best used in tandem with other, more conventional interventions. Due to the limited quality and quantity of included studies, large-scale and well-designed randomized controlled trials are warranted.

***(J Med Internet Res 2022;24(9):e37618)*** doi:10.2196/37618

**KEYWORDS**

spinal cord injury; pressure injury, telemedicine; systematic reviews; meta-analyses; network meta-analyses; review; spinal injury; spinal cord; pressure injury; injury

## Introduction

Spinal cord injury (SCI) is a disabling and costly disease, the incidence of which is increasing year by year. The incidence of SCI is estimated to be between 12 and 65 cases per million globally [1] and between 13 and 60 cases per million in China [2]. More than 20% of patients with SCI develop pressure injury as a result of motor and sensory dysfunction, limited body movement, or the long period of time spent in a bed or wheelchair [3]. The daily cost of pressure injury treatment per adult patient ranges from €1.71 (US $1.70) to as much as €470 (US $468.31), and the cost of treating severe pressure injury is even higher [4]. In addition, 7%-8% of deaths among patients with SCI are directly attributable to pressure injury [5]. Pressure injury seriously affects the quality of life of patients with SCI and places a heavy care burden and economic burden on their families and society [6].

There are well-understood measures that can reduce the incidence of pressure injury, and prevention is more cost-effective than treatment [7]. However, most countries have insufficient medical resources, and particularly insufficient professional expertise in the community [8], to provide the necessary long-term and multidisciplinary follow-up of community-dwelling patients with SCI. That results in many obstacles to preventing and treating pressure injury. Today, however, it is becoming more feasible to provide medical services including diagnosis and information about self-care remotely through a variety of communication technologies, including video consultation via mobile apps [9,10]. This has been applied to the prevention and treatment of pressure injury among community-dwelling patients with SCI, but its effectiveness and safety remains inadequately confirmed because systematic studies vary in their sample sizes and conclusions.

Until now, there has been no systematic review of the applicability of telemedicine in preventing and treating pressure injury among community-dwelling patients with SCI. That motivated this systematic review and network meta-analysis. Network meta-analysis can assess both direct and indirect evidence [11]. The biggest advantage is allowing for the simultaneous inclusion of multiple pairwise comparisons in a series of different interventions and ranking the interventions. Therefore, systematic review and network meta-analysis was used to evaluated the effectiveness of telemedicine on the prevention and treatment of pressure injury among community-dwelling patients with SCI, and determine which telemedicine form is more effective, to provided evidence useful for clinical practice.

## Methods

### Study Design and Search Strategy

This systematic review and network meta-analysis was performed according to the PRISMA-NMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Network Meta-Analysis) standards [12]. It was registered in the PROSPERO database (International Prospective Register of Systematic Reviews; ID: CRD42020194061).

The databases searched were the China National Key Information corpus, Wanfang, CBM, VIP, Embase, PubMed, Cochrane Library, Web of Science, Scopus, and ProQuest. The dates searched were from establishment of each database to September 30, 2021. Multimedia Appendix 1 shows the keywords used to search each corpus. Keywords and search strategy were designed by the first author, then reviewed by a librarian. Other clinical trial registration websites (Science-paper Online, Open Grey, ClinicalTrials.gov, and China’s Clinical Trial Registry) were searched manually, and references to related papers and reviews were followed up.

Only randomized controlled trials (RCTs) and quasi-experimental studies were included in the systematic review, and only RCTs were included in the network meta-analysis. Beyond that, 4 other criteria were applied.

1. Participants: community-dwelling persons with an SCI.
2. Interventions: complete or partial telemedicine intervention. In complete telemedicine intervention, there was no face-to-face contact during the trial, only telemedicine intervention by telephone, video, or mobile app. Treatment involving only one form of telemedicine intervention was designated as a single complete telemedicine intervention, while therapy combining two or more forms of telemedicine intervention was called a mixed complete telemedicine intervention. Partial telemedicine intervention designated treatment combining telemedicine with a nontelemedicine intervention (such as an outpatient follow-up visit or a home visit).
3. Controls: The “no telemedicine” cases included nontelemedicine intervention and also health guidance only before discharge treated as a blank control. A second type of control was where there was another group treated differently from the experimental group, such as when the experimental group used video and the control group used the telephone. A third case was self-control studies with no control group.
4. Outcomes: Primary and secondary outcomes were considered. The primary outcomes were the incidence of pressure injury, the rate of healing of the pressure injury, and pressure injury severity (size, depth, and Pressure Ulcer Scale for Healing [PUSH]). Any economic data reported were treated as a secondary outcome.

Certain reports had to be excluded, for example, academic meeting abstracts or papers without full text; papers published repeatedly; and papers for which adequate data could not be obtained even after contacting the authors.

### Data Extraction

Two authors (the first and the second author) worked independently and blindly to screen titles, abstracts, and full texts, and select studies applying the inclusion and exclusion criteria. Any disagreements were resolved by discussion or by consulting the corresponding author. EndNote X9 software (Clarivate) was first used to exclude duplicates. Then, reading the title and abstract was enough to exclude clearly irrelevant papers. Finally, reading the full text allowed us to determine whether or not a study should be included. If necessary, authors were contacted by email or telephone for further information.

The first and second authors also worked independently and blindly to extract data and assess the risk of bias, again consulting the corresponding author if necessary. The data extracted included each study's characteristics, participant characteristics, intervention and control treatments, and outcomes. The Cochrane risk of bias tool [13] was used to assess the RCTs, and the Joanna Briggs Institute critical appraisal tool [14] was used with the quasi-experimental studies.

### Statistical Analysis

$I^2$ statistic was used to evaluate the consistency of the results of included studies, with 25%, 50%, and 75% representing low, moderate, and high heterogeneity, respectively [15]. The fixed-effect model was used when the heterogeneity was acceptable ($I^2 \leq 50\%$, $P>.10$), otherwise the random effect model was used. If the heterogeneity was still too large after subgroup analysis or sensitivity analysis, if the number of studies was too small, or if the data could not be synthesized, only descriptive analysis was performed. Dichotomous data were analyzed using relative risk (RR) and 95% CIs. Continuous data were analyzed using weighted mean difference (WMD) where the same tools were used, and standardized mean difference where different studies used different tools. When $P<.05$, the difference between the two groups was statistically significant.

Song [16] has proposed that network meta-analysis should satisfy hypotheses about homogeneity, similarity, and consistency. Otherwise, the nonconformity needs to be explained, or network meta-analysis should not be performed. The evaluation and treatment of the homogeneity requirement is the same as with the heterogeneity of traditional direct comparison meta-analysis. There is no recognized statistical test for verifying the similarity hypothesis, so it must be evaluated based on the characteristics of the included studies. The inconsistency model, node splitting, and inconsistency factors are commonly used to evaluate consistency, with $P>.05$ indicating good consistency in a closed loop if the 95% CI starting point of the inconsistency factor was 0, indicating that the direct and indirect evidence was consistent. RR, WMD, and standardized mean difference were also computed with their 95% CIs. The surface under the cumulative ranking (SUCRA) curve was used to calculate the ranking probabilities of the treatments. The SUCRA values range from 0% to 100%, and the higher the value, the better the result. Comparison-adjusted funnel plots were also used to assess the potential for small study effects.

## Results

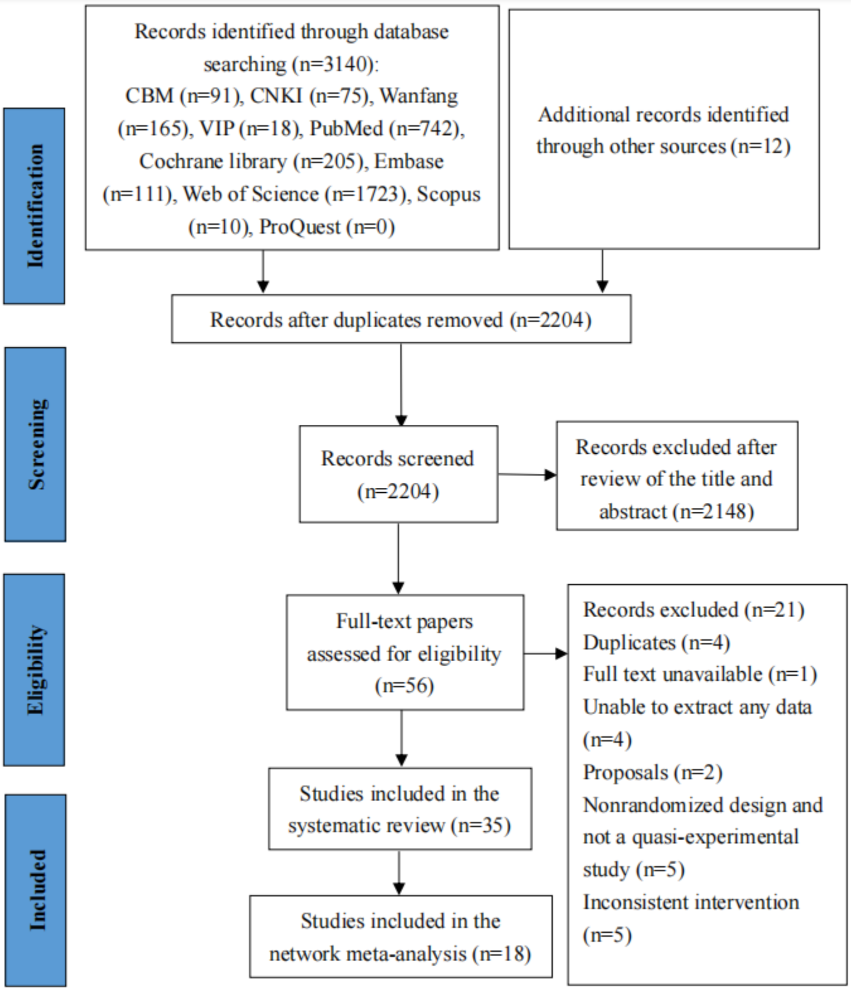
### Study Selection

The search found 3152 studies. Of those, 948 duplicates were excluded through EndNote. Reading the titles and abstracts of 2204 reports led to 2148 being excluded as irrelevant. Finally, 56 studies were screened in full text, of which 21 were excluded and 35 were finally included (Figure 1).

The 35 studies included 25 RCTs [7,17-40] and 10 quasi-experimental studies [41-50] (Multimedia Appendix 2). The 3131 subjects were community-dwelling patients with SCI, aged 18-96 years old. Of 3131 patients, 2226 (71.1%) were male. Trauma was the most common cause of injury. The top specific causes were traffic accidents, fall from height, fall with heavy objects, and other fall. The study durations ranged from 1 week to 2 years. In most of the studies, nurses served as the main researcher managing diet and nutrition, elimination, and pressure injury, and preventing other complications. They also provided related education and guidance.

Mixed complete telemedicine interventions mainly used the WeChat app and telephone. The average utilization was about 1 hour every day to answer questions, once weekly to convey relevant knowledge, and perhaps a weekly face-to-face video chat if necessary. Telephone calls were made on average once per month, when needed. Single complete telemedicine interventions were mainly delivered via telephone. The average frequency was about once per week. Partial telemedicine intervention was usually a combination of telephone or video telemedicine with outpatient follow-up or home visits. The frequency was about once per week by telephone or face-to-face video, and once per month for outpatient follow-up or home visit. The main form of nontelemedicine intervention was outpatient follow-up or home visit. The frequency was about once per month. The blank control group only received health education before discharge, but the patients could call a medical professional when they needed help.

**Figure 1.** Flowchart for search and selection of the included studies.



## Bias Assessment

The overall quality of the studies included was categorized as acceptable. Approximately half of the studies reported randomization, but some reports lacked details about any allocation blinding, which could cause potential selection bias. No study was judged as “low risk” in terms of performance bias because it is very difficult to blind patients in telemedicine intervention trials. About one-quarter of the studies blinded the outcome assessors. There was no evidence of attrition bias, selective reporting bias, and other bias in any of the included studies (Multimedia Appendix 3). In the quasi-experimental studies, item 3 was judged to be “not applicable” to 4 studies [41,46,47,50] because they were self-controlled. Item 7 was judged as “unclear” in 4 studies [44,45,48,49]. The other items all received a “yes” (see Multimedia Appendix 4 for details).

## Meta-analysis and Descriptive Analysis Results

Overall, 27 studies [17-21,23-29,31-35,37-40,42-45,48,49] reported the incidence of pressure injury among their community-dwelling subjects. Among the studies, 18 were RCTs [17-21,24-29,31-35,37,38] analyzed by network meta-analysis, 5 were quasi-experimental studies [42-44,48,49] analyzed by meta-analysis, and 4 [23,39,40,45] could not be combined for descriptive analysis.

The meta-analysis showed that the incidence of pressure injury was significantly lower in the telemedicine intervention group (n=468; RR 0.24, 95% CI 0.14-0.41; *P*<.05; $I^2$=0%, fixed-effects model; Multimedia Appendix 5). The other 4 studies which could not be combined also found that the incidence of pressure injury in the intervention group was lower than in the control group (*P*<.05).

A total of 9 studies [17,21,22,41,44,46,47,49,50] reported the rate of pressure injury healing. There were 3 RCTs [17,21,22] and 6 quasi-experimental studies [41,44,46,47,49,50]. Due to the limited sample size, 4 studies [41,46,47,50] (self-controlled) were classified as the telemedicine intervention group. The other 5 [17,21,22,44,49] were descriptive because the data could not be combined. The meta-analysis showed that the rate of pressure injury healing was significantly faster in the telemedicine intervention group (n=55; RR 0.73, 95% CI 0.62-0.85; *P*<.05; $I^2$=0%, fixed-effects model; Figure 2). In the other 5 studies without meta-analysis, 14 patients in the telemedicine intervention group healed, along with 9 patients in the control group.

A total of 4 RCTs [7,19,27,36] reported on the severity of the pressure injury studied. The meta-analysis showed that patients in a telemedicine intervention group tended to have lower PUSH scores (n=162; WMD=–1.98, 95% CI –3.51 to –0.46; *P*=.01;

$I^2$=0%, fixed-effects model; Figure 3). Although one study [7] found no significant difference in the improvement of pressure injury area and depth, the improvement in the other telemedicine groups was significantly better than in the corresponding control group.

Only one study (an RCT) [7] reported economic data. The incremental cost-effectiveness ratio was 2306 Indian rupees (approximately US $130) per 1 $cm^2$ reduction in pressure injury area and 44,915 Indian rupees (US $2523) per additional quality-adjusted life year. This result shows that the telemedicine intervention was cost-effective, at least in India.

**Figure 2.** The effectiveness of telemedicine on the healing rate of pressure injury. ES: effect size.

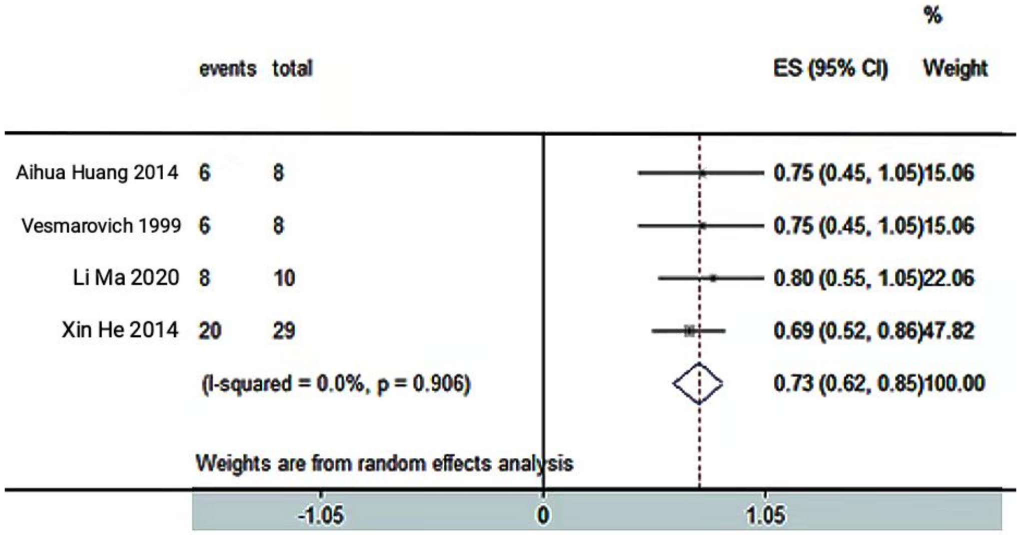
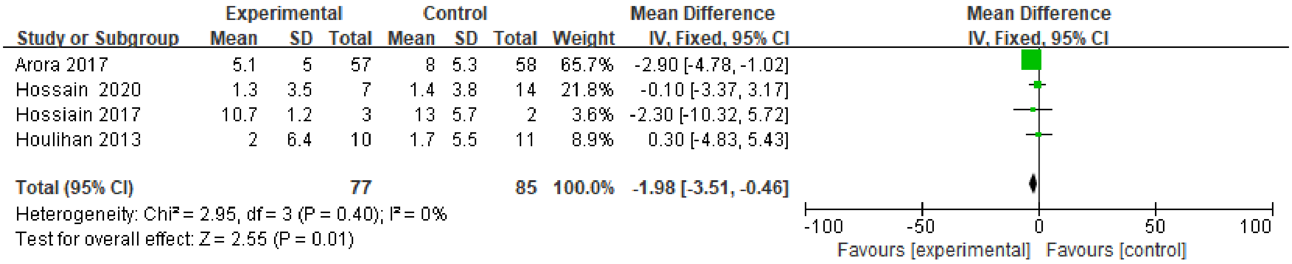


**Figure 3.** Comparison of the effectiveness of telemedicine and control on PUSH scores. PUSH: Pressure Ulcer Scale for Healing.



## Network Meta-analysis Results

A total of 18 RCTs [17-21,24-26,28,29,31-35,37,38,45] involving 5 forms of intervention were included in the network meta-analysis. A network plot for the incidence of pressure injury was produced using the STATA software package (version 14.0; StataCorp LLC). Nodes indicated treatments, with the size of each node proportional to the number of observations in the sample. The thickness of the lines was proportional to the number of studies directly comparing a pair of treatments. Two triangles were formed with the interventions in this study: triangle 134 and triangle 135. The most numerous comparable studies involved partial telemedicine intervention with a blank control (Figure 4).

A consistency test did not identify statistically significant inconsistency ($X^2$=3.76, $P$=.15). The loops were consistent, since their 95% CI included 0. Node splitting showed no statistically significant difference between the direct and indirect estimate of the summary effect (see Multimedia Appendix 6 for details).

The SUCRA estimates (Figure 5) and the SUCRA value (Multimedia Appendix 7) show that mixed complete telemedicine intervention was the best form of intervention for reducing the incidence of pressure injury. Mixed complete telemedicine intervention (93.5%) was better than partial telemedicine intervention (80.5%), which was better than nontelemedicine intervention (32.7%), single complete telemedicine intervention (31.7%), and blank control (11.7%) (see Table 1 for details). The comparison-adjusted funnel plot was basically symmetrical, indicating that the possibility of publication bias was small (Multimedia Appendix 8).

**Figure 4.** Network meta-analysis of eligible comparisons for incidence. 1: blank control; 2: nontelemedicine intervention; 3: single complete telemedicine intervention; 4: mixed complete telemedicine intervention; 5: partial telemedicine intervention.

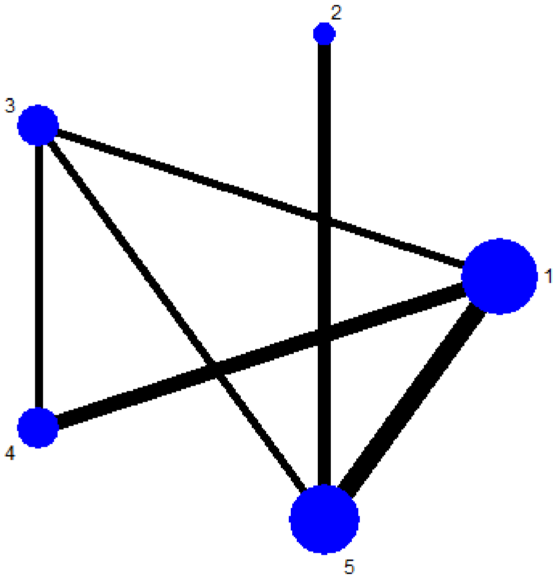


**Figure 5.** The surface under the cumulative ranking estimate. 1: blank control; 2: nontelemedicine intervention; 3: single complete telemedicine intervention; 4: mixed complete telemedicine intervention; 5: partial telemedicine intervention.

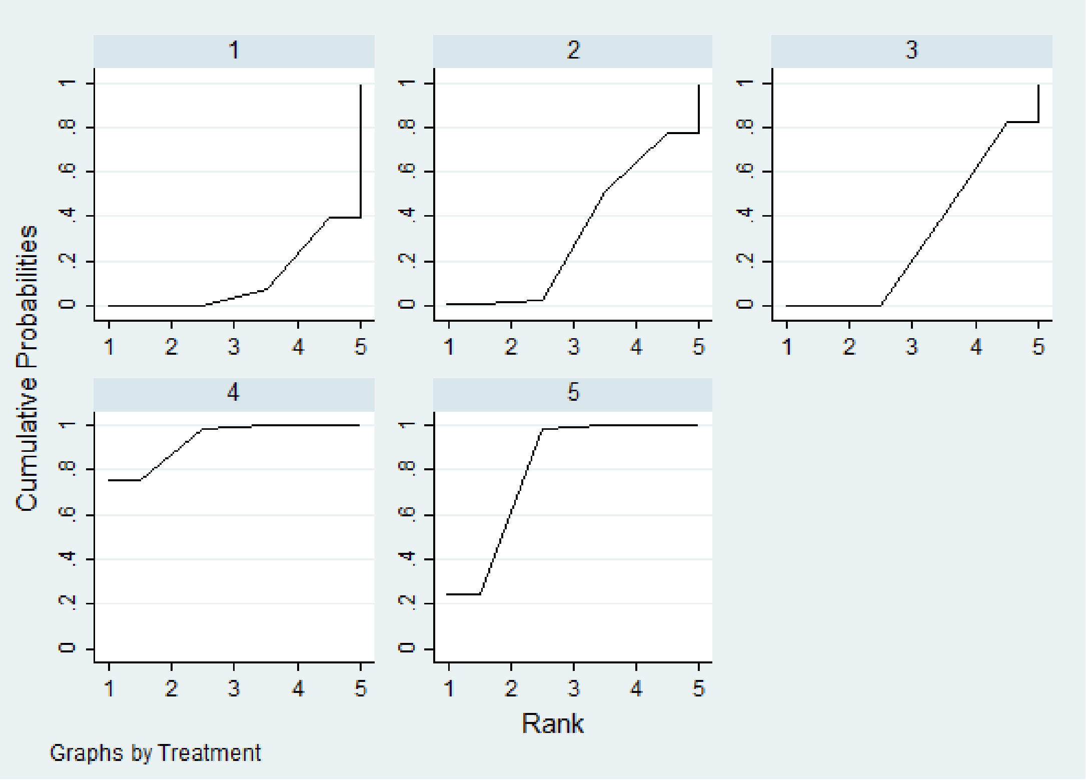


**Table 1.** The effectiveness of telemedicine in preventing pressure injury according to the network meta-analysis.

| MCTI[a] | PTI[b] | SCTI[c] | NTI[d] | Blank control |
|---|---|---|---|---|
| MCTI | 1.34 (0.59-3.06) | 3.52 (1.95-6.36) | 3.35 (1.07-10.48) | 4.63 (2.16-9.93) |
| 0.75 (0.33-1.70) | PTI | 2.62 (1.29-5.36) | 2.50 (1.13-5.49) | 3.45 (2.09-5.71) |
| 0.28 (0.16-0.51) | 0.38 (0.19-0.78) | SCTI | 0.95 (0.33-2.75) | 1.31 (0.67-2.59) |
| 0.30 (0.10-0.93) | 0.40 (0.18-0.88) | 1.05 (0.36-3.04) | NTI | 1.38 (0.54-3.52) |
| 0.22 (0.10-0.46) | 0.29 (0.18-0.48) | 0.76 (0.39-1.50) | 0.72 (0.28-1.84) | Blank control |

[a] MCTI: mixed complete telemedicine intervention.

[b] PTI: partial telemedicine intervention.

[c] SCTI: single complete telemedicine intervention.

[d] NTI: nontelemedicine intervention.

## Discussion

### Principal Findings

This systematic review and network meta-analysis results show that telemedicine intervention can reduce the incidence and severity of pressure injury and improve the rate of healing of such injuries without increasing the medical economic burden on community-dwelling patients with SCI. In addition, the results indicate that combining telemedicine with conventional interventions is the most effective form of intervention for preventing pressure injury.

Although the overall quality of the studies was regarded as acceptable, none were able to blind the participants and personnel. Some reports mentioned random sequence generation, allocation concealment, and blinding, but without specifics. That may relate to the space limitations of journal publications or the design of the experiments. This review included only studies published in Chinese or English, of which many were Chinese. This may be related to the Quality Nursing Service demonstration project launched by China's National Health Commission in early 2010 [51] and to its "Internet+" policy implemented in 2016 [52]. More than 10 studies in this systematic review were conducted using QQ and WeChat, which are Chinese social media platforms that support sending text, pictures, and videos, and support multiperson group chats via the internet. Such social media software, which have a high penetration rate, broad mass base, rich features, and no extra charge, are likely to be central to the future development of telemedicine.

The results show that telemedicine intervention can reduce the incidence and severity of pressure injury. As part of rehabilitation, patients with SCI were usually educated in preventive skin care techniques, but they are often not continued after discharge [53,54]. Contacting former patients in the community through telemedicine can improve compliance [36], but the prevention and treatment of pressure injury after an SCI involves several medical disciplines. In addition, it is also necessary to pay attention to any motor or sensory dysfunction, self-care ability limitations, and nutritional status after discharge [55,56]. Carlson's study [23] has shown that telemedical support from a multidisciplinary team can provide rehabilitation, nutrition suggestions, and psychological guidance as well as how to deal with the threat of pressure injury. When a discharged patient has health-related problems, they can get appropriate help in time. Effort should be devoted in clinical practice to promoting multidisciplinary team cooperation and comprehensively promoting the physical and mental recovery of patients with SCI.

The hospital stays of patients with SCI are shorter now than in the past [57,58]. That allows less time for patients to receive education, rehabilitation, and adjustment, making them more likely to benefit from subsequent telemedicine. In Vesmarovich's study [50], the patients and their families were given guidance on dressing techniques before discharge, and video was used to give continued medical care after discharge. That improved the rate of pressure injury healing. Many primary care doctors lack the expertise and skills to deal with the complex needs of patients with SCI [59], but access to specialized rehabilitation institutions is costly and might be difficult to arrange. When patients have insufficient resources to cope with the disease, they are more likely to aggravate pressure injury [60]. Huang [46] reports using a combination of telemedicine and nontelemedicine techniques to help medical staff change pressure injury dressings during home visits. Families were trained by telephone. Home visits allow for face-to-face treatment of pressure injury and providing professional guidance. They can to some extent compensate for the reductions in education time caused by shorter hospital stays.

The results show that using telemedicine did not increase the economic burden of SCI. Most developed countries provide patients with SCI with any equipment they may need to cope with their injury. They receive training before discharge and then remote written or oral guidance without the need for professionals to enter the patient's home. That helps to minimize the cost of an SCI [40]. A study by Xu [61] showed that telemedicine can save money without reducing efficacy. The studies included in this systematic review rarely discuss cost considerations. In most of them, the patients received any necessary equipment for free or at low cost. Future research should conduct a rigorous cost-benefit analysis to demonstrate not only the impact on patient health, but also the value of investing in telemedicine intervention.

The network meta-analysis showed that the best intervention for preventing pressure injury combined two or more forms of telemedicine. The most common combination was internet chat (usually WeChat) with telephone conversations. Patients and their carers cannot be assumed able to identify pressure injury early and take countermeasures soon enough of their own accord [40]. Professionals, though, can observe patients' skin using pictures or video and provide timely medical advice, thereby reducing the incidence of pressure injury. At present, the diagnostic accuracy from using images compared with that achieved through face-to-face evaluation remains unclear. That needs further documentation in well-designed studies with large samples. Of course, even if patients and carers receive the knowledge they need, over time, that knowledge may well be gradually forgotten. In China, the typical caregiver is older than their patient. More than 30% of caregivers are over 60 years old [62]. Moreover, even some middle-aged persons cannot effectively use a smartphone and a networking platform. They prefer telephoning or outpatient follow-up [63]. Nevertheless, repeated instruction, whether by telephone or internet messaging, and regular push messages to remind caregivers, can effectively reduce the incidence of pressure injury [64]. Any telemedicine intervention should of course suit the individual patient's needs, condition, home situation, and level of medical understanding.

## Limitations

This study was to some unknown extent restricted by being limited to reports in either Chinese or English. Beyond that, some experimental studies were not included because they were unfinished or the relevant data could not be extracted. That may induce a certain degree of publication bias. There were also reports that did not describe the intervention frequency in detail, and some in which the accuracy of individual outcomes was relatively low due to the small number of related studies. More high-quality RCTs with large samples are need for further demonstration.

## Conclusions

Current evidence shows that telemedicine is an economical and feasible form of intervention. It can reduce the incidence of pressure injury in community-dwelling patients with SCI. Combining telemedicine with other sorts of intervention is better than using telemedicine alone. Telemedicine can improve the rate of pressure injury healing and reduce the severity of the injury without increasing the medical economic burden on patients with SCI. These above conclusions need to be further verified by additional high-quality RCTs using large samples. Future studies could explore the research on telemedicine in languages other than Chinese and English.

## Acknowledgments

This study was supported by the Natural Science Foundation of Guangdong Province (grant 2021A1515011800). The authors thank to Dr Yingchun Zeng of Hong Kong Polytechnic University for her help in statistical analysis.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search details.
[DOCX File , 12 KB - jmir_v24i9e37618_app1.docx ]

Multimedia Appendix 2
Characteristics of the studies included.
[PDF File (Adobe PDF File), 2511 KB - jmir_v24i9e37618_app2.pdf ]

Multimedia Appendix 3
Risk of bias assessment summary for each Cochrane item.
[DOCX File , 170 KB - jmir_v24i9e37618_app3.docx ]

Multimedia Appendix 4
Risk of bias assessment for the quasi-experimental studies.
[DOCX File , 13 KB - jmir_v24i9e37618_app4.docx ]

Multimedia Appendix 5
The effectiveness of telemedicine on the incidence of pressure injury (quasi-experimental studies).
[DOCX File , 69 KB - jmir_v24i9e37618_app5.docx ]

Multimedia Appendix 6
Node splitting results.
[DOCX File , 12 KB - jmir_v24i9e37618_app6.docx ]

Multimedia Appendix 7
Relative treatment rankings.
[DOCX File , 11 KB - jmir_v24i9e37618_app7.docx ]

---

Multimedia Appendix 8
A comparison-adjusted funnel plot of the studies.
[DOCX File , 38 KB - jmir_v24i9e37618_app8.docx ]

---

## References

1. Hamid R, Averbeck MA, Chiang H, Garcia A, Al Mousa RT, Oh S, et al. Epidemiology and pathophysiology of neurogenic bladder after spinal cord injury. World J Urol 2018 Oct;36(10):1517-1527. [doi: 10.1007/s00345-018-2301-z] [Medline: 29752515]
2. Chen X, Chen D, Chen C, Wang K, Tang L, Li Z. The epidemiology and disease burden of traumatic spinal cord injury in China: A systematic review. Chin J Evid-Based Med 2018 Jan 1;18:143-150. [doi: 10.7507/1672-2531.201712034]
3. Chen H, Cai J, Du L, Shen H, Yu H, Song Y, et al. Incidence of pressure injury in individuals with spinal cord injury: A systematic review and meta-analysis. J Wound Ostomy Continence Nurs 2020;47(3):215-223. [doi: 10.1097/WON.0000000000000633] [Medline: 32384524]
4. Demarré L, Van Lancker A, Van Hecke A, Verhaeghe S, Grypdonck M, Lemey J, et al. The cost of prevention and treatment of pressure ulcers: A systematic review. Int J Nurs Stud 2015 Nov;52(11):1754-1774. [doi: 10.1016/j.ijnurstu.2015.06.006] [Medline: 26231383]
5. Scovil CY, Ranabhat MK, Craighead IB, Wee J. Follow-up study of spinal cord injured patients after discharge from inpatient rehabilitation in Nepal in 2007. Spinal Cord 2012 Mar;50(3):232-237. [doi: 10.1038/sc.2011.119] [Medline: 22025245]
6. Irgens I, Midelfart-Hoff J, Jelnes R, Alexander M, Stanghelle JK, Thoresen M, et al. Videoconferencing in pressure injury: randomized controlled telemedicine trial in patients with spinal cord injury. JMIR Form Res 2022 Apr 19;6(4):e27692 [FREE Full text] [doi: 10.2196/27692] [Medline: 35438645]
7. Arora M, Harvey LA, Glinsky JV, Chhabra HS, Hossain S, Arumugam N, et al. Telephone-based management of pressure ulcers in people with spinal cord injury in low- and middle-income countries: a randomised controlled trial. Spinal Cord 2017 Mar;55(2):141-147. [doi: 10.1038/sc.2016.163] [Medline: 27995939]
8. Zakrasek EC, Creasey G, Crew JD. Pressure ulcers in people with spinal cord injury in developing nations. Spinal Cord 2015 Jan;53(1):7-13. [doi: 10.1038/sc.2014.179] [Medline: 25366536]
9. Irgens I, Kleven L, Sørli H, Stanghelle JK, Rekand T. Telemedicine brings specialist healthcare services to patients' homes. Tidsskr Nor Laegeforen 2015 Oct 20;135(19):1716-1717 [FREE Full text] [doi: 10.4045/tidsskr.15.0770] [Medline: 26486663]
10. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR mHealth uHealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
11. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002 Aug 30;21(16):2313-2324. [doi: 10.1002/sim.1201] [Medline: 12210616]
12. Zhang X, Tan R, Lam WC, Cheng CW, Yao L, Wang X, et al. PRISMA extension for moxibustion 2020: recommendations, explanation, and elaboration. Syst Rev 2020 Oct 25;9(1):247 [FREE Full text] [doi: 10.1186/s13643-020-01502-7] [Medline: 33100229]
13. 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]
14. Zhou Y, Gu Y, Hu Y, Xing W. JBI evidence-based health center's quality assessment tool for different types of research: quality evaluation of intervention research. J Nurs Train 2018 Jan 22:112-113. [doi: 10.16821/j.cnki.hsjx.2018.01.007]
15. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
16. Song F, Loke YK, Walsh T, Glenny A, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009 Apr 03;338:b1147 [FREE Full text] [doi: 10.1136/bmj.b1147] [Medline: 19346285]
17. Li Q, Li J, Pan H. Effects of Online Home Nursing Care Model Application on Patients with Traumatic Spinal Cord Injury. Risk Manag Healthc Policy 2021;14:1703-1709 [FREE Full text] [doi: 10.2147/RMHP.S301874] [Medline: 33935524]
18. Li J, Li Q, Yang B. Participatory continuous nursing using the WeChat platform for patients with spinal cord injuries. J Int Med Res 2021 May;49(5):3000605211016145 [FREE Full text] [doi: 10.1177/03000605211016145] [Medline: 34038208]
19. Hossain MS, Harvey LA, Islam MS, Rahman MA, Muldoon S, Biering-Sorensen F, et al. A community-based intervention to prevent serious complications and death 2 years after discharge in people with spinal cord injury in Bangladesh (CIVIC):

a randomised trial. Spinal Cord 2021 Jun;59(6):649-658 [FREE Full text] [doi: 10.1038/s41393-020-00546-9] [Medline: 32917948]

20. Zou S, Mo X. Study on the application of continuous nursing based on Wechat platform in discharged patients with lumbar fracture with spinal cord injury. World Latest Medicine Information 2020 Jan 1;20(54):41-42. [doi: 10.3969/j.issn.1671-3141.2020.54.023]
21. Chen Y. The influence of continuous nursing method on the quality of life and rehabilitation of patients with spinal cord injury. J Clin Nurs Pract 2020 Mar 25;5(13):27-28 [FREE Full text]
22. Kryger MA, Crytzer TM, Fairman A, Quinby EJ, Karavolis M, Pramana G, et al. The effect of the interactive mobile health and rehabilitation system on health and psychosocial outcomes in spinal cord injury: Randomized controlled trial. J Med Internet Res 2019 Aug 28;21(8):e14305 [FREE Full text] [doi: 10.2196/14305] [Medline: 31464189]
23. Carlson M, Vigen CL, Rubayi S, Blanche EI, Blanchard J, Atkins M, et al. Lifestyle intervention for adults with spinal cord injury: Results of the USC-RLANRC Pressure Ulcer Prevention Study. J Spinal Cord Med 2019 Jan;42(1):2-19 [FREE Full text] [doi: 10.1080/10790268.2017.1313931] [Medline: 28414254]
24. Dai X, Yuan Y. Effect of continuous nursing through WeChat platform on postoperative rehabilitation of patients with cervical spinal cord injury and paraplegia. Clin Res Pract 2019 Jul 11;4(20):145-147. [doi: 10.19347/j.cnki.2096-1413.201920061]
25. Zhao K. Effect of continuous nursing intervention on postoperative rehabilitation of patients with spinal fracture complicated with spinal cord injury. Chin J Trauma Disabil Med 2018 Mar 01;26(3):76-78. [doi: 10.13214/j.cnki.cjotadm.2018.03.052]
26. Li J, Gao D, Pan P, Huang R, Huang J. Effect of continuous nursing on rehabilitation and quality of life in patients with spinal cord injury. Chin Med Pharm 2018 Apr 11;8(8):137-139. [doi: 10.3969/j.issn.2095-0616.2018.08.043]
27. Hossain M, Harvey L, Rahman M, Bowden J, Islam M, Taylor V, et al. A pilot randomised trial of community-based care following discharge from hospital with a recent spinal cord injury in Bangladesh. Clin Rehabil 2017 Jun;31(6):781-789. [doi: 10.1177/0269215516654207] [Medline: 27311454]
28. Yang Y, Zhang J, Yang J. Comparison of the effects of regular outpatient review and continuous nursing after discharge on patients with spinal cord injury. J Clin Med Pract 2017 Feb 10;21(14):44-47. [doi: 10.7619/jcmp.20171401]
29. Li C, Hu J, Zheng L, Su X, Hui H, Yu J. Application of family members' participatory continuous nursing based on Wechat platform in postoperative rehabilitation of patients with cervical spinal cord injury. Chin J Mod Nurs 2017 Dec 26;23(36):4599-4602. [doi: 10.3760/cma.j.issn.1674-2907.2017.36.009]
30. Ma L, Yan L, Guoyan W, Jing W, Zhengfeng Y, Li W. The effect of Wechat follow-up and telephone follow-up in home nursing care of patients with spinal cord injury. Chin J Clin Rational Drug Use 2017 May 20;10(14):159-160. [doi: 10.15887/j.cnki.13-1389/r.201714.113]
31. Luo X, Guo N, Xu H, Li Y. Effect of continuous nursing intervention on postoperative rehabilitation of patients with spinal fracture complicated with spinal cord injury. Hebei Med J 2017 Aug 30;39(16):2542-2544. [doi: 10.3969/j.issn.1002-7386.2017.16.041]
32. Jing W. Application of continuous nursing in prevention of pressure injury among patients with spinal fracture complicated with paraplegia after discharge. Chin Med Guide 2017 Sep 30;15(27):208-209. [doi: 10.15912/j.cnki.gocm.2017.27.170]
33. Feng Y. Effect of WeChat intervention on the prognosis of hip decubitus. Chin J Prim Med Phar 2017 Sep 18;24(18):2875-2876. [doi: 10.3760/cma.j.issn.1008-6706.2017.18.040]
34. Fan X. The effect of the WeChat joint nursing outpatient follow-up on the quality of patients with spinal cord injury. Int J Nurs 2017 Dec 30;36(24):3330-3333. [doi: 10.3760/cma.j.issn.1673-4351.2017.24.005]
35. Yang Y. Analysis of the preventive effect of continuous nursing on pressure ulcer among patients with spinal fracture complicated with paraplegia after discharge. J Chin Three Gorges Univ(Natural Sci) 2017 Jun 27;39(S1):258-259 [FREE Full text]
36. Houlihan BV, Jette A, Friedman RH, Paasche-Orlow M, Ni P, Wierbicky J, et al. A pilot study of a telehealth intervention for persons with spinal cord dysfunction. Spinal Cord 2013 Sep;51(9):715-720. [doi: 10.1038/sc.2013.45] [Medline: 23752260]
37. Ru Z, Wu H. Effect of community rehabilitation guidance on complications and quality of life in patients with Spinal Cord injury. Chin Rural Health Serv Adm 2013 Nov 20;33(11):1231-1233 [FREE Full text]
38. Fan X. Effect of follow-up intervention on rehabilitation of patients with spinal cord injury. Contemporary Med 2011 Nov 15;17(32):118-120. [doi: 10.3969/j.issn.1009-4393.2011.32.083]
39. Dallolio L, Menarini M, China S, Ventura M, Stainthorpe A, Soopramanien A, THRIVE Project. Functional and clinical outcomes of telemedicine in patients with spinal cord injury. Arch Phys Med Rehabil 2008 Dec;89(12):2332-2341. [doi: 10.1016/j.apmr.2008.06.012] [Medline: 19061746]
40. Phillips VL, Temkin A, Vesmarovich S, Burns R, Idleman L. Using telehealth interventions to prevent pressure ulcers in newly injured spinal cord injury patients post-discharge. Results from a pilot study. Int J Technol Assess Health Care 1999 Oct 01;15(4):749-766. [doi: 10.1017/s0266462399154l4x]
41. Ma L, Wu Y, Long Y, Wen J, Yang Z, Wang L. Application of Wechat follow-up in community nursing. J Clin Nurs Sci 2020 Jan 15;5(3):5-6 [FREE Full text]

42. Chen Q, Wu H, Li N. Effect of continuous nursing on physical and mental health of patients with complete thoracic spinal cord injury after discharge. Chin J Prac Nurs 2017 Jun 11;33(17):1281-1286. [doi: 10.3760/cma.j.issn.1672-7088.2017.17.001]
43. Li J. Effect of continuous health education on family function and health behavior of patients with thoracolumbar fracture and spinal cord injury.: Wenzhou Medical University; 2017. URL: https://d.wanfangdata.com.cn/thesis/D01293876 [accessed 2017-06-02]
44. Xu A. Influence of continuity of care on rehabilitation effect of spinal cord injury patients. Chin Nurs Res 2016 Jun 1;30(16):2038-2040. [doi: 10.3969/j.issn.1009-6493.2016.16.039]
45. Wang Y, Sun F, Sun Y, Zhou Y. Application of continuous nursing in prevention of pressure ulcer among patients with spinal fracture complicated with paraplegia after discharge. Chin Military Med J Southeast 2015 Mar 20;2:198-199. [doi: 10.3969/j.issn.1672-271X.2015.02.028]
46. Huang A, Lu L, He J. Effects of home-based nursing interventions for patients with spinal cord injury. Nurs J Chin PLA 2014 Jan 15;31(01):35-37. [doi: 10.3969/j.issn.1008-9993.2014.01.010]
47. He X, Feng Y, Hu C. Application of continuous nursing in pressure ulcer patients with high paraplegia after discharge. Hainan Med J 2014 Sep 10;17:2647-2648. [doi: 10.3969/j.issn.1003-6350.2014.17.1040]
48. Xia T, Hou G, Jiang Z, Luo X, Duan L, Li Y. Effect of family follow-up nursing on complications in community-dwelling patients with traumatic paraplegia. J Nurs Train 2012 May 22;27(10):936-938. [doi: 10.3969/j.issn.1002-6975.2012.10.032]
49. Lian F, Meng Y. The effect of regular intensive rehabilitation on guidance on discharged paraplegia patients. J Hebei Med Univ 2010 Aug 31;31(8):915-917. [doi: 10.3969/j.issn.1007-3205.2010.08.012]
50. Vesmarovich S, Walker T, Hauber RP, Temkin A, Burns R. Use of telerehabilitation to manage pressure ulcers in persons with spinal cord injuries. Adv Wound Care 1999 Jun;12(5):264-269. [Medline: 10655800]
51. Guo Y. Review of nursing management in 2009 and prospect in 2010. Chin Nurs Manage 2010 Jan 15;1(10):5-6. [doi: 10.3969/j.issn.1672-1756.2010.01.001]
52. Integrated development plan of informatization and industrialization (2016-2020). Ministry of Industry and Information Technology. Beijing: Ministry of Industry and Information Technology; 2016 Oct 12. URL: https://www.ndrc.gov.cn/fggz/fzzlgh/gjjzxgh/201706/W020191104624315735098.doc [accessed 2016-10-12]
53. Hug K, Stumm C, Debecker I, Fellinghauer CS, Peter C, Hund-Georgiadis M. Self-Efficacy and Pressure Ulcer Prevention After Spinal Cord Injury-Results From a Nationwide Community Survey in Switzerland (SwiSCI). PM R 2018 Jun;10(6):573-586. [doi: 10.1016/j.pmrj.2017.11.017] [Medline: 29225161]
54. Yang Y, Chang G, Hsu M, Chang J. Remote monitoring of sitting behaviors for community-dwelling manual wheelchair users with spinal cord injury. Spinal Cord 2009 Jan;47(1):67-71. [doi: 10.1038/sc.2008.72] [Medline: 18560373]
55. De Miguel-Rubio A, Rubio MD, Alba-Rueda A, Salazar A, Moral-Munoz JA, Lucena-Anton D. Virtual reality systems for upper limb motor function recovery in patients with spinal cord injury: systematic review and meta-analysis. JMIR mHealth uHealth 2020 Dec 03;8(12):e22537 [FREE Full text] [doi: 10.2196/22537] [Medline: 33270040]
56. Vos-Draper TL, Morrow MMB. Seating-Related pressure injury prevention in spinal cord injury: A review of compensatory technologies to improve In-Seat movement behavior. Curr Phys Med Rehabil Rep 2016 Dec;4(4):320-328 [FREE Full text] [doi: 10.1007/s40141-016-0140-7] [Medline: 28603664]
57. Abou-Amsha K, Front L, Gelernter I, Hart J, Catz A. [Rehabilitation outcomes in patients with spinal cord injury 1992-2003: survival, neurologic recovery, length of stay in hospital]. Harefuah 2008 Jun;147(6):504-8, 575. [Medline: 18693626]
58. Mahmoud H, Qannam H, Zbogar D, Mortenson B. Spinal cord injury rehabilitation in Riyadh, Saudi Arabia: time to rehabilitation admission, length of stay and functional independence. Spinal Cord 2017 May;55(5):509-514 [FREE Full text] [doi: 10.1038/sc.2016.165] [Medline: 28139661]
59. Rosin N, Tabibi R, Trimbath J, Henzel M. A primary care provider's guide to prevention and management of pressure injury and skin breakdown in people with spinal cord injury. Top Spinal Cord Inj Rehabil 2020;26(3):177-185 [FREE Full text] [doi: 10.46292/scI2603-177] [Medline: 33192045]
60. Floríndez LI, Carlson ME, Pyatak E, Blanchard J, Cogan AM, Sleight AG, et al. A qualitative analysis of pressure injury development among medically underserved adults with spinal cord injury. Disabil Rehabil 2020 Jul;42(15):2093-2099 [FREE Full text] [doi: 10.1080/09638288.2018.1552328] [Medline: 30621460]
61. Xu C, Jackson M, Scuffham PA, Wootton R, Simpson P, Whitty J, et al. A randomized controlled trial of an interactive voice response telephone system and specialist nurse support for childhood asthma management. J Asthma 2010 Sep;47(7):768-773. [doi: 10.3109/02770903.2010.493966] [Medline: 20716014]
62. Chang F, Xie H, Yang Y, Zhang J, Wang H, Wu A. Analysis on burden of care and influencing factors of caregivers for patients with spinal cord injury in community. Shanghai Nurs J 2019 Dec 21;12(19):20-24. [doi: 10.3969/j.issn.1009-8399.2019.12.005]
63. Chen Y, Ma F, Tian S. Investigation and analysis of the needs of patients with spinal cord injury for continuous nursing services of chest and lung physical therapy. Chin General Pract Nurs 2016 Sep 10;25(14):2599-2601. [doi: 10.3969/j.issn.1674-4748.2016.25.003]
64. Groah SL, Schladen M, Pineda CG, Hsieh CJ. Prevention of pressure ulcers among people with spinal cord injury: A systematic review. PM R 2015 Jun;7(6):613-636. [doi: 10.1016/j.pmrj.2014.11.014] [Medline: 25529614]

## Abbreviations

**PRISMA-NMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Network Meta-Analysis
**PROSPERO:** International Prospective Register of Systematic Reviews
**PUSH:** Pressure Ulcer Scale for Healing
**RCT:** randomized controlled trial
**RR:** relative risk
**SCI:** spinal cord injury
**SUCRA:** surface under the cumulative ranking
**WMD:** weighted mean difference

*Edited by G Eysenbach; submitted 28.02.22; peer-reviewed by B Smith, S Bhattacharjya, RL Kirby; comments to author 21.05.22; revised version received 02.07.22; accepted 09.08.22; published 07.09.22.*

*Please cite as:*
*Chen G, Wang T, Zhong L, He X, Huang C, Wang Y, Li K*
*Telemedicine for Preventing and Treating Pressure Injury After Spinal Cord Injury: Systematic Review and Meta-analysis*
*J Med Internet Res 2022;24(9):e37618*
*URL: https://www.jmir.org/2022/9/e37618*
*doi:10.2196/37618*
*PMID:36069842*

Original Paper

# Enrollment and Retention of Participants in Remote Digital Health Studies: Scoping Review and Framework Proposal

Paola Daniore[1,2], MSc; Vasileios Nittas[3], PhD; Viktor von Wyl[1,2,3], Prof Dr

[1]Institute for Implementation Science in Healthcare, University of Zurich, Zurich, Switzerland
[2]Digital Society Initiative, University of Zurich, Zurich, Switzerland
[3]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

**Corresponding Author:**
Viktor von Wyl, Prof Dr
Institute for Implementation Science in Healthcare
University of Zurich
Universitatstrasse 84
Zurich, 8006
Switzerland
Phone: 41 0446343762
Email: viktor.vonwyl@uzh.ch

## Abstract

**Background:** Digital technologies are increasingly used in health research to collect real-world data from wider populations. A new wave of digital health studies relies primarily on digital technologies to conduct research entirely remotely. Remote digital health studies hold promise to significant cost and time advantages over traditional, in-person studies. However, such studies have been reported to typically suffer from participant attrition, the sources for which are still largely understudied.

**Objective:** To contribute to future remote digital health study planning, we present a conceptual framework and hypotheses for study enrollment and completion. The framework introduces 3 participation criteria that impact remote digital health study outcomes: (1) participant motivation profile and incentives or nudges, (2) participant task complexity, and (3) scientific requirements. The goal of this study is to inform the planning and implementation of remote digital health studies from a person-centered perspective.

**Methods:** We conducted a scoping review to collect information on participation in remote digital health studies, focusing on methodological aspects that impact participant enrollment and retention. Comprehensive searches were conducted on the PubMed, CINAHL, and Web of Science databases, and additional sources were included in our study from citation searching. We included digital health studies that were fully conducted remotely, included information on at least one of the framework criteria during recruitment, onboarding or retention phases of the studies, and included study enrollment or completion outcomes. Qualitative analyses were performed to synthesize the findings from the included studies.

**Results:** We report qualitative findings from 37 included studies that reveal high values of achieved median participant enrollment based on target sample size calculations, 128% (IQR 100%-234%), and median study completion, 48% (IQR 35%-76%). Increased median study completion is observed for studies that provided incentives or nudges to extrinsically motivated participants (62%, IQR 43%-78%). Reducing task complexity for participants in the absence of incentives or nudges did not improve median study enrollment (103%, IQR 102%-370%) or completion (43%, IQR 22%-60%) in observational studies, in comparison to interventional studies that provided more incentives or nudges (median study completion rate of 55%, IQR 38%-79%). Furthermore, there were inconsistencies in measures of completion across the assessed remote digital health studies, where only around half of the studies with completion measures (14/27, 52%) were based on participant retention throughout the study period.

**Conclusions:** Few studies reported on participatory factors and study outcomes in a consistent manner, which may have limited the evidence base for our study. Our assessment may also have suffered from publication bias or unrepresentative study samples due to an observed preference for participants with digital literacy skills in digital health studies. Nevertheless, we find that future remote digital health study planning can benefit from targeting specific participant profiles, providing incentives and nudges, and reducing study complexity to improve study outcomes.

***(J Med Internet Res 2022;24(9):e39910)*** doi:10.2196/39910

**KEYWORDS**

remote digital health studies; remote clinical trials; remote cohorts; digital epidemiology; digital health; health outcome; conceptual framework; user-centered design; population-based digital health; participant recruitment; interventional study

## Introduction

### Background

The widespread availability of smartphones (estimated to be 3.6 billion users worldwide [1]) presents the opportunity to involve diverse population groups in health research. Mobile technologies, such as smartphones and wearables, have come to play a central role in health research, giving rise to digital health studies that are conducted partly or entirely remotely. Although there is no unified definition, we define remote digital health studies as longitudinal studies that use mobile technologies to conduct all key steps of a study completely online [2]. Remote digital health studies promise significant cost, time, and scalability advantages when compared with traditional studies, by allowing key steps of the study investigations to be conducted in real-time and without in-person presence [3,4]. Overcoming the barriers of time and physical presence, remote digital health studies allow for the long-term monitoring of larger populations and thus promise to advance health research and patient care delivery [5-7].

Despite these opportunities, recent studies report high participant attrition rates, likely partially attributable to the lack of in-person interactions between researchers and study participants. Other studies highlight the risk of recruitment bias, especially with younger, more affluent, and often healthier populations being overrepresented in studies with digital technologies [8-10]. These concerns point toward a possible imbalance between participants who typically join remote digital health studies and participants who are often underrepresented, but may benefit the most from remote digital health research and monitoring. This may be a result from a lack of understanding of the motivators, facilitators, and barriers that enable participation in remote digital health studies [11,12].

Trends of participant enrollment and retention have been widely investigated in traditional research settings [13-15] and with digital health studies following Eysenbach's Law of Attrition [16]. However, there is a paucity of evidence that supports study planning in remote digital health research. Furthermore, most study planning recommendations for remote digital health studies are based on qualitative methods and focus on scientific, rather than participant-specific requirements [17,18]. This presents an unmet need for quantitative, evidence-based guidance that informs remote digital health study planning to enable high enrollment and retention through a person-centered lens.

### Aims

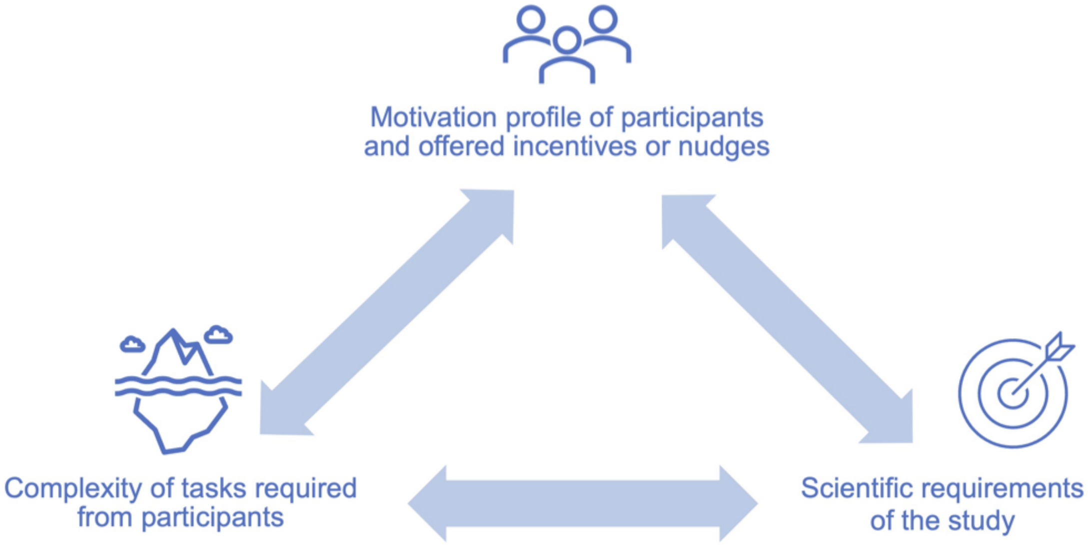
This review aims to explore participant enrollment and retention for remote digital health studies. We introduce a framework on the interplay between 3 criteria explored and validated in previous digital health studies: (1) participant motivation profile and incentives or nudges, (2) participant task complexity, and (3) scientific requirements. We propose hypotheses and explore them with a scoping review of remote digital health studies specifically, focusing on methodological aspects that affect participant enrollment and retention. The goal of this scoping review is to inform the planning and implementation of remote digital health studies from a person-centered perspective.

### Conceptual Framework and Hypotheses

We introduce a conceptual framework that encompasses the main factors that affect digital health study enrollment and retention from a person-centered lens, to guide the extraction of relevant information. The framework is based on the notion that enrollment and retention in remote digital health studies are influenced by 3 elements: (1) participant motivation profile and incentives or nudges, (2) participant task complexity, and (3) scientific requirements (Figure 1). Our conceptual framework and hypotheses were informed by Eysenbach [16], previous large-scale remote digital health studies [5,19], as well as by our personal experiences in planning and conducting 2 longitudinal remote digital health studies [20,21]. A more detailed description of the framework development process can be found in Multimedia Appendix 1 [2,5,14,16,19-28].

We define hypotheses to explore in this study. Specifically, we expect that incentives and nudges increase participant motivation to enroll in and complete a study (Hypothesis 1). On the contrary, we expect a decrease in enrollment and study completion with increased complexity of study tasks (Hypothesis 2a). Finally, we also expect that participants in interventional studies may be willing to endure higher task complexity than participants enrolled in observational studies (Hypothesis 2b).

**Figure 1.** Guiding framework for remote digital health studies.



## Conceptual Framework Definitions

### *Motivation Profile of Participants and Offered Incentives*

Participants' motivation to enroll and complete remote digital health studies can be *intrinsic* or *extrinsic*. Intrinsic participation may be motivated by altruistic motives or by hopes for medical advances, especially among individuals with chronic diseases. Extrinsic participation can be motivated, among others, by investigators in the form of monetary incentives or clinical referrals [29,30]. The provision of incentives or nudges may help extrinsically motivated participants to enroll and participate in a study. Incentives are predominantly offered in the form of monetary compensation, while nudges mainly come in the form of reminders or personal contact [31-34]. Intrinsically motivated participants do not necessarily require incentives or nudges to enroll or participate in a study. It is to be noted, however, that intrinsic and extrinsic motivations are not necessarily mutually exclusive from one another and may coincide in a study. In this review, we conduct an exploratory assessment based on the aforesaid definitions. Different approaches to assessing participant motivations may exist.

### *Complexity of Tasks Required From Participants*

The *frequency* and *complexity* of tasks required from the participants, along with the expected *duration of the study*, can impact study completion. Tasks can be categorized as physical and mental tasks. Typical physical tasks required of participants include physical activity tasks, such as walking a certain number of steps every day, as well as other essential tasks to fulfill the study's requirements, such as signing an informed consent form, filling out questionnaires at baseline or at follow-up, or logging of health-related outcomes [27,35,36]. Mental tasks, such as the measure of cognitive burden of a participant, are harder to quantify as they typically rely on patient-reported outcome measures [37,38]. In digital heath studies, task complexity is compounded by electronic measurements and device handling, which may require high digital literacy skills [39-41]. Task complexity can be reduced by accompanying measures, such as passive data collection, or technical support. Applied to remote digital health studies, high-complexity tasks come in the form of many required tasks over long study periods. Provision of assistance during participant onboarding may improve digital literacy skills of participants, while also instilling a sense of trust between the participant and the researcher through personal interactions [42,43].

### *Scientific Requirements of the Study*

Scientific requirements define the *study design* and expected *target sample size*. Therefore, scientific requirements set the goal of the study, while task complexity or incentives or nudges enable study goal achievement. Target sample sizes are generally estimated through statistical power analyses for enrollment goals [27,44]. Consideration of participant requirements may help increase the statistical power and reduce selection bias of the study. However, sample size calculation methods often may not anticipate participant losses to follow-up and failure to complete tasks [45]. Sensitivity analyses may be conducted after the study to assess the impact of deviations in participant enrollment or retention.

# *Methods*

## Study Outcome Definitions

In our study we refer to 3 phases of participants' involvement in remote digital health studies: recruitment, onboarding, and retention (Tables 1 and 2). We summarize the outcomes of the 3 phases of participant involvement in remote digital health studies as study enrollment and completion. A detailed approach for defining each phase of the study and the outcomes can be found in Multimedia Appendix 1.

The 3 criteria in our proposed framework can affect all phases of remote digital health studies. In the next sections, we will explore our hypotheses by assessing the outcomes of interest for each framework criteria across all 3 phases of remote digital health studies.

**Table 1.** Phases of digital health studies.

| Phases | Definition |
|---|---|
| Recruitment | Fulfillment of requirements for study enrollment |
| Onboarding | Provision of (technical) assistance to start study tasks |
| Retention | Fulfillment of requirements for study completion |

**Table 2.** Outcomes of digital health studies.

| Outcomes | Measure |
|---|---|
| Study enrollment target | (Achieved enrollment/target enrollment) × 100% |
| Study completion | Percentage of enrolled participants who completed the study |

### Search Strategy and Study Selection

To explore our hypotheses, we conducted a scoping review according to the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-analyses Extension for Scoping Reviews) checklist [46] (Multimedia Appendix 2). Our search was performed on the PubMed, CINAHL, and Web of Science databases for primary research articles published between January 1, 2016, and June 31, 2021. We limited our search to this period based on the results of a preliminary search revealing a paucity of remote digital health studies published before 2016 [19,47]. We also assessed reviews and included relevant primary studies for a full-text review based on citation searches. The complete search strategy for each database can be found in Multimedia Appendix 3.

Our selection was guided by the criteria outlined in Textbox 1. Screening was conducted in 2 phases. Initially, we screened titles and abstracts and then the full texts. For both phases, the entire screening was conducted by one investigator (PD), while a second investigator (VN) performed checks on a randomly selected sample of studies in the title and abstract screening (80/662 articles, 12.1%) and in the full-text screening (50/150, 33.3%). Any disagreements were discussed and, if required, consensus was achieved through the third investigator (VvW). Agreement was 75/80 articles (94%) for title and abstract screening and 45/50 articles (90%) for full-text screening.

**Textbox 1.** Literature inclusion criteria.

- Studies that match the definition of a remote digital health study (ie, digital health study where all steps are conducted online and without in-person interactions between participants and study investigators).
- Studies that mention their approach to recruit, onboard, or retain participants.
- Studies that mention approaches to at least one of the proposed framework criteria.
- Studies that provide evidence on study enrollment or study completion.

### Data Extraction and Synthesis

Data extraction was standardized yet developed iteratively. In cases where research articles referenced to the original protocol for the same study, the additional protocols were assessed to collect missing information of interest. The initial data extraction was based on standard study characteristics (eg, study design, participant characteristics) and guided by the conceptual framework (Figure 1). During the full-text screening, other criteria of interest (eg, measure of study completion) were identified as relevant and retrospectively included.

We conducted qualitative analyses to explore our hypotheses. Qualitative analyses are presented for the entire study sample, as well as for samples stratified based on the median study duration. Descriptions of qualitative data were provided to summarize key findings from the included studies within the structure of the conceptual framework. We also conducted an exploratory quantitative assessment of the framework criteria with the study enrollment and completion outcomes. The correlations between measures relevant to our study's framework criteria and the study enrollment and completion outcomes retrieved from the included studies were assessed using Spearman rank correlation for continuous variables and the Kruskal-Wallis rank-sum test for categorical variables.

All screening and extraction procedures were completed in MS Excel (Microsoft, Inc.). All statistical analyses were completed in R, version 4.0.0 (R Foundation for Statistical Computing) using ggplot2, version 3.3.3, for plots. The threshold for statistical significance was set at $P<.05$ (2-tailed testing).

### Data Availability

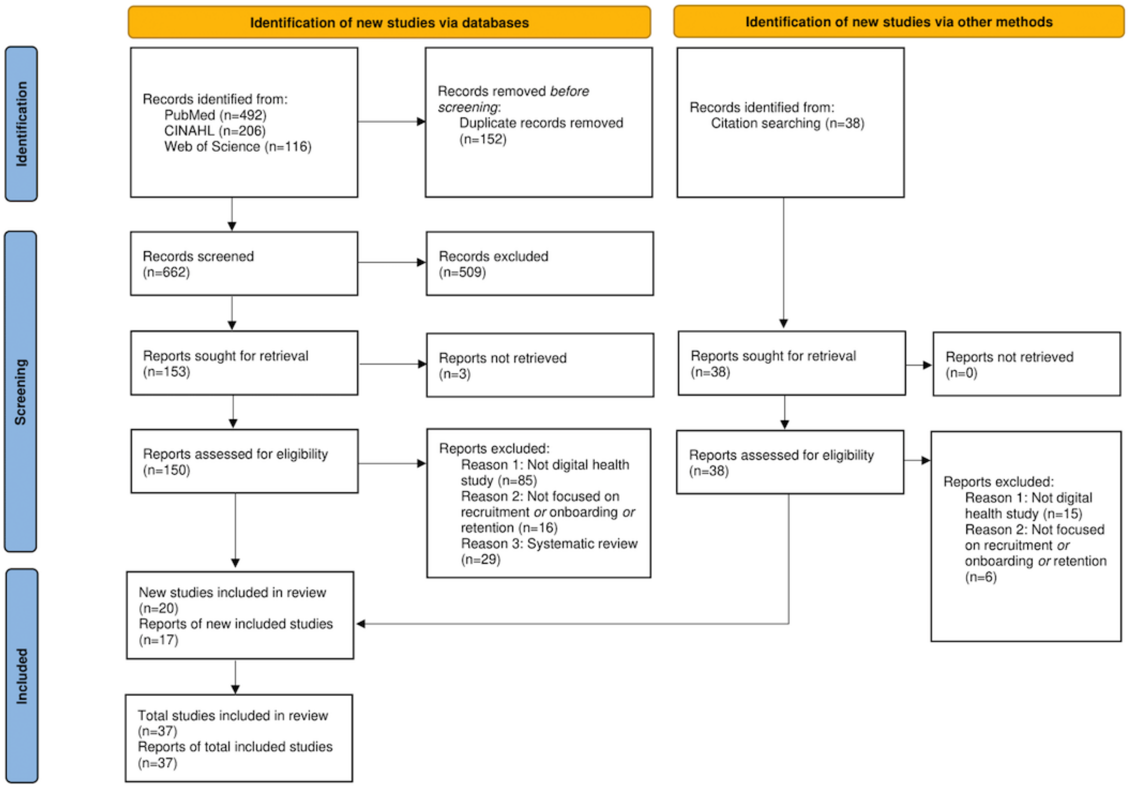
The data from the papers that support the findings of this study are publicly available. All data used in this review can be found in Multimedia Appendices 4-6 [6,7,43,48-81].

## Results

### Data Search and Identification of Articles

Our database searches yielded 662 references (Figure 2). Twenty studies fulfilled all inclusion criteria and were finally included. An additional 38 studies were identified through citation searches in relevant systematic reviews, of which 17 met all inclusion criteria. Overall, 37 studies were included for assessment in this review.

**Figure 2.** Flowchart for study identification, screening, and inclusion.



## General Description of Included Studies

The study characteristics, study requirements, and framework criteria outcomes of the final study sample are presented in Multimedia Appendix 1 and summarized in Multimedia Appendices 4-6 [6,7,43,48-81], respectively. Of the 37 included studies, 19/37 (51%) were randomized control trials, 11/37 (30%) nonexperimental studies, and 7/37 (19%) quasi-experimental studies. All studies were longitudinal, 28/37 (76%) of which were interventional and 9/37 (24%) observational studies.

The most prevalent therapeutic areas in our study sample were psychiatry (11/37 studies, 30%), neurology (4/37 studies, 11%), and addiction (4/37 studies, 11%). Most studies (12/37, 32%) measured physical activity levels, 3/37 (8%) measured smoking cessation, 3/37 (8%) measured depression management, 2/37 (5%) measured stress management, 2/37 (5%) measured pain management, and the rest of the studies (15/37, 41%) measured other outcomes. Most studies relied on smartphones for the study procedure (23/37, 62%) and predominantly measured step counts (7/37, 19%). Most studies (34/37, 92%) collected patient-reported outcomes.

Most studies (25/37, 68%) targeted participants with an existing health condition, mainly enrolled females (median female enrollment of 77%, IQR 52%-88%) and participants with a median age of 39 years (IQR 35-47 years). A median of 43% (IQR 32-49%) of enrolled participants had a lower educational background, while information on enrolled participant income (27/37, 73%) and employment status (25/37, 68%) was largely missing from the studies.

## Description of Included Studies According to Framework

A description of the included studies based on our framework's criteria is found in Table 3. The table summarizes the participants' motivation profiles, the most prevalent incentives and nudges provided at each stage of the study (more information is available in Multimedia Appendix 6), the frequencies of the tasks for the studies and measures provided to reduce participant burden, the target sample size of the studies, and the study outcomes. Given the scarcity of mental task measures in the assessed studies, our review only included measures of physical tasks. The table is stratified by the median duration of the included studies (12 weeks, IQR 12-26 weeks).

A total of 27 studies reported on study completion (Multimedia Appendix 5); 14/27 (52%) studies assessed for completion based on participant retention throughout the study period, 10/27 (37%) studies assessed for completion based on researcher-defined metrics (eg, completion of 1 task within a 30-day period) and 3/27 (11%) studies assessed for completion based on the fulfillment of all study tasks.

In the next sections, study enrollment and completion are assessed based on our study's hypotheses. An assessment of the task frequencies and study durations is made against measures provided to reduce participant burden, as well as incentives or nudges provided to participants. Further descriptions and correlations between the studies' framework criteria and outcomes are reported in Multimedia Appendix 7.

**Table 3.** Summary statistics of included studies based on framework criteria.

| Profile | Study duration[a]: ≤12 weeks (n=19) | Study duration: >12 weeks (n=15) | Study duration: unknown duration (n=3) | Overall (n=37) |
|---|---|---|---|---|
| **Motivation profile of participants and offered incentives or nudges** | | | | |
| **Motivation profiles, n (%)** | | | | |
| Extrinsic | 14 (74) | 8 (53) | 0 (0) | 22 (59) |
| Intrinsic | 5 (26) | 7 (47) | 3 (100) | 15 (41) |
| **Incentives/nudges: recruitment[b], n (%)** | | | | |
| Monetary | 4 (21) | 3 (20) | 2 (67) | 9 (24) |
| Referral source | 3 (16) | 1 (7) | 0 (0) | 4 (11) |
| Vested interest | 0 (0) | 2 (13) | 1 (33) | 3 (8) |
| **Incentives/nudges: onboarding, n (%)** | | | | |
| Personal assistance | 1 (5) | 5 (33) | 0 (0) | 6 (16) |
| Peer support | 0 (0) | 1 (7) | 0 (0) | 1 (3) |
| **Incentives/nudges: retention, n (%)** | | | | |
| Monetary | 10 (53) | 6 (40) | 1 (33) | 15 (41) |
| Reminders | 10 (53) | 5 (33) | 0 (0) | 15 (41) |
| Personal contact | 9 (47) | 3 (20) | 0 (0) | 12 (32) |
| Incentives/nudges: minimum one phase[c], n (%) | 18 (95) | 13 (87) | 2 (67) | 33 (89) |
| **Complexity of tasks required from participants** | | | | |
| **Task complexity: study tasks** | | | | |
| Monthly steps, median (IQR)[d] | 16 (5-30) | 30 (29-36) | —[e] | 28 (8-31) |
| Total steps, median (IQR)[f] | 35 (12-93) | 99 (14-180) | 3.00 (3.0-3.0) | 58 (10-120) |
| **Task complexity: reduction of burden, n (%)** | | | | |
| Passive monitoring | 7 (37) | 9 (60) | 0 (0) | 16 (43) |
| Short, repetitive daily tasks | 6 (32) | 8 (53) | 0 (0) | 14 (38) |
| At least one burden reduction[g] | 9 (47) | 12 (80) | 0 (0) | 21 (57) |
| **Scientific requirements of the study** | | | | |
| Target sample size, median (IQR) | 72 (50-120) | 313 (238-838) | 473 (336-609) | 200 (50-350) |
| **Study outcomes** | | | | |
| Enrolled participants, median (IQR) | 281 (89-450) | 560 (150-2800) | 100 (55-200) | 300 (89-950) |
| Retained participants, median (IQR) | 110 (45-240) | 800 (190-1700) | — | 180 (70-690) |
| Enrollment target (%), median (IQR) | 150 (124-302) | 101 (96-125) | 82 (48-116) | 128 (100-234) |
| Study completion (%), median (IQR) | 48 (38-73) | 55 (32-79) | — | 48 (35-76) |

[a] The duration of the study as defined in the study protocol.

[b] The top 3 recruitment and retention incentives and nudges are reported; more information is available in Multimedia Appendix 6.

[c] Incentive or nudge provided in at least one of the study phases: recruitment, onboarding, or retention.

[d] The number of physical tasks investigators required participants to do on a monthly basis throughout the study duration as defined in the study protocol.

[e] Not available.

[f] The total number of physical tasks investigators required participants to do throughout the study duration as defined in the study protocol.

[g] Burden reduction for participants either through 1 of the 2 approaches, passive monitoring or short, repetitive daily tasks, provided in a study.

## Evidence for Hypotheses

### *Hypothesis 1: Study Outcomes Based on Motivation Profiles and Incentives or Nudges*

In this section we link evidence from the studies' *task complexity* with the *participant motivation profile and incentives or nudges* criteria of our framework to assess study enrollment and completion outcomes. The 15 studies that enrolled intrinsically motivated participants reached a median enrollment target of 137% (IQR 98%-226%), and the 22 studies that enrolled extrinsically motivated participants reached a median enrollment target of 126% (IQR 102%-213%). The median completion rate of studies that focused on intrinsically motivated participants was 41% (IQR 20%-49%), whereas those that included extrinsically motivated participants had a median study completion of 62% (IQR 43%-78%; Figure 3).

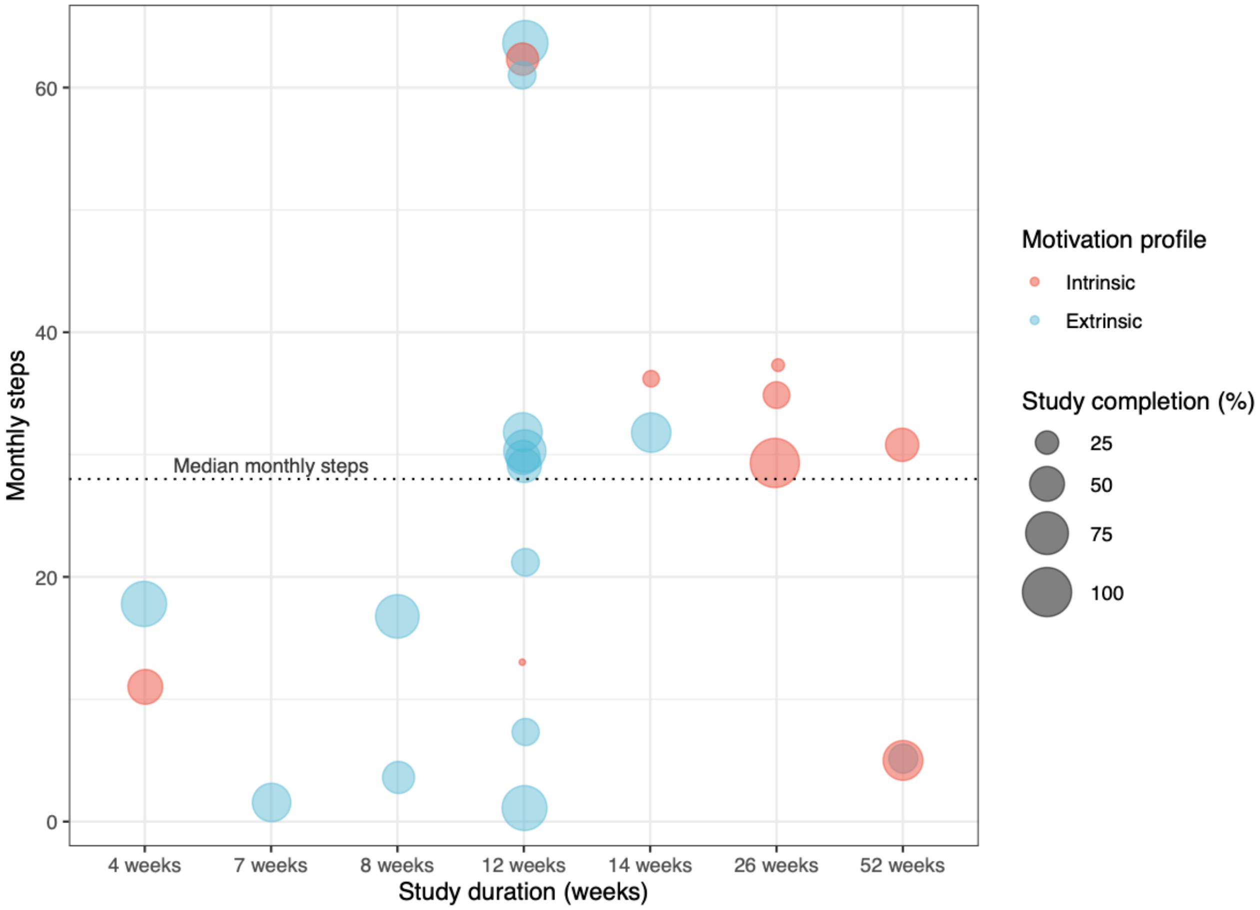
Studies that enrolled intrinsically motivated participants had a median duration of 14 weeks (IQR 12-26) and had participants complete a median of 30 (IQR 12-36) tasks per month. Approximately half of these studies (7/15, 47%) offered passive monitoring of health data or had participants complete lower complexity, repetitive daily tasks (7/15, 47%).

Studies that enrolled extrinsically motivated participants had a median duration of 12 weeks (IQR 12-14) and had participants complete a median of 26 (IQR 7-30) tasks per month. Some of the studies (9/22, 41%) offered passive monitoring of health data and fewer offered participants lower complexity, repetitive daily tasks (7/22, 32%).

Incentives or nudges for recruitment of intrinsically motivated participants were offered only in 4/15 (27%) studies and incentives or nudges for retention were offered in 9/15 (60%) studies. For studies that targeted extrinsically motivated participants, incentives or nudges for recruitment were offered in 12/22 (55%) studies and incentives or nudges for retention were offered in all (22/22, 100%) studies.

Statistical significance testing did not reveal evidence for an effect of intrinsic or extrinsic study motivations or the provision of incentives and nudges on study enrollment or completion outcomes (Multimedia Appendix 7).

**Figure 3.** Study completion based on participant motivations and study requirements.



### *Hypothesis 2: Study Outcomes Based on Task Complexity and Study Design*

In this section we link evidence from the study *task complexity* with the *scientific requirements* criteria of our framework to assess study enrollment and completion outcomes.

Interventional studies achieved a median enrollment target of 128% (IQR 100%-150%) and observational studies a median enrollment target of 103% (IQR 102%-370%). Interventional studies achieved a median study completion of 55% (IQR 38%-79%) and observational studies achieved a median study completion of 43% (IQR 22%-60%; Figure 4).

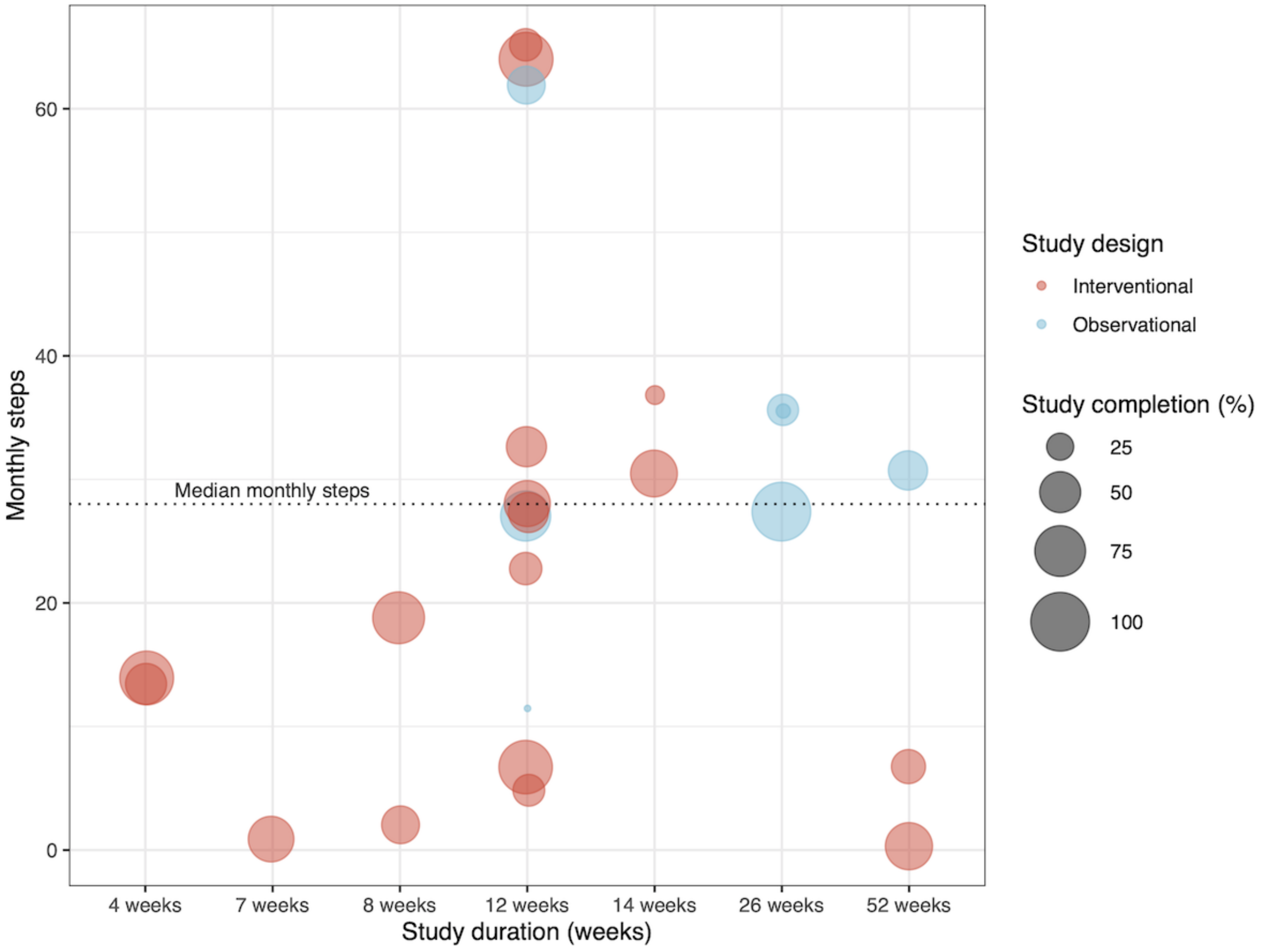
A total of 28 interventional studies had a median duration of 12 weeks (IQR 12-14) and had participants complete a median of 16 (IQR 4-30) tasks per month. Few interventional studies (8/28, 29%) offered passive monitoring of health data and had participants complete lower complexity, repetitive daily tasks (8/28, 29%).

Nine observational studies had a median duration of 26 weeks (IQR 12-26 weeks) and required participants to complete a median of 30 (IQR 29-35) tasks per month. Most observational studies (7/9, 78%) offered passive monitoring of health data and had participants complete lower complexity, repetitive daily tasks (6/9, 67%).

Approximately half of the interventional studies (15/28, 54%) provided incentives or nudges for recruitment, and 26 interventional studies (26/28, 93%) provided incentives or nudges for retention. Only 1 observational study (1/9, 11%) provided an incentive for recruitment, while 5 (5/9, 56%) of the observational studies provided incentives or nudges for participant retention.

Statistical analyses did not reveal evidence for an effect of task complexity or study design on study enrollment or completion outcomes (Multimedia Appendix 7). The observed coefficients only suggest weak or no correlations, which were not statistically significant (Multimedia Appendix 7).

**Figure 4.** Study completion based on study design and requirements.



## Discussion

### Principal Findings

The studies from our sample were primarily interventional. Most studies targeted extrinsically motivated participants for shorter-length studies and provided incentives or nudges to recruit or retain participants. Around half of the studies provided measures to reduce participant burden through passive data collection or by requiring participants to complete frequent, shorter tasks. The study sample revealed high values of achieved target participant enrollment and retention. These findings suggest that the criteria defined in our framework may have an impact on the studies' enrollment and completion outcomes.

Exploring our first hypothesis, high study enrollment is observed in studies that targeted intrinsically motivated participants. Study completion in these studies was lower than in those that targeted extrinsically motivated participants. This may be due to researchers requiring intrinsically motivated participants to complete more tasks per month in comparison to studies with extrinsically motivated participants or due the provision of additional incentives in studies with extrinsically motivated participants. Furthermore, the provision of lower complexity, daily repetitive tasks was similar between studies with intrinsically or extrinsically motivated participants. Our study's preliminary qualitative analyses, although the results did not reach statistical significance, indicate that the provision of incentives or nudges as well as the reduction of required tasks from participants may contribute to higher study completion.

Exploring our second hypothesis, differences in study enrollment and completion outcomes are also observed between interventional and observational study designs. Here, higher enrollment is observed in interventional studies, despite higher efforts from most of the observational studies in our sample to achieve wider reach study recruitment and simplify enrollment procedures. Interventional studies revealed a higher study completion, although fewer measures to reduce participant burden were provided. We find that our framework provides preliminary relevant criteria and findings relevant to remote digital health study participation to guide researchers in study planning.

## Comparison With Prior Work

### *Descriptive Assessment of Participation Strategies in Remote Digital Health Studies*

Descriptive results from our study sample reveal a preference for investigators to provide incentives and nudges to retain rather than recruit participants. Researchers' choice to focus on retaining participants could be in response to reported high attrition rates in digital health studies [15,83-85]. If participants are not retained, the quality of the evidence base of a study is decreased. From our study sample, the lower provision of incentives or nudges at recruitment could have likely been motivated by higher participant reach through open social networks, multimodal recruitment strategies, and user-friendly interfaces [48-64,80,81]. The replacement of in-person enrollment procedures with mobile-forward procedures that are easy to use, specifically, enables easier study enrollment, as mentioned in 8 of our studies [6,50,53,55-57,65,66]. However, it is important to consider that successful recruitment strategies do not necessarily translate into high participant retention. This could be due, in part, to unrepresentative samples that enroll in a study to collect financial incentives and then dropout [86,87]. Studies may be highly effective if they place equal importance on their recruitment and retention strategies, while applying sample validation approaches to ensure the representativeness of their study sample.

Studies of shorter durations reported lower study completion than those of longer durations. This is inconsistent with the observation that studies that ran for 12 weeks or less required less tasks to be completed by participants than in studies' that ran for over 12 weeks. However, half of the longer duration studies reduced participant burden by asking them to complete repetitive, short-length tasks once a day [52,54,56,62,67-69,80]. For most of these studies, these tasks could be completed at any point in time during the day and the completion of all tasks was made optional. The majority of longer duration studies also collected data passively through smartphones or wearable sensors [7,48,51,54,60,61,67,70,71]. These findings are in line with recent literature on best practices to maximize participation in longitudinal digital health studies [88,89]. Specifically, evidence aligns with our observations that the integration of short, repetitive tasks that align with participants' daily lives is more likely to achieve higher participation in comparison to longer, infrequent tasks with higher cognitive burden [49,72,90].

Few studies in our sample involved onboarding procedures for participants. This may be due to investigators' possible preference for in-person interactions during study enrollment. Previous literature on this topic suggests that interacting with participants virtually, rather than in person, may contribute to higher study attrition [16,83,91]. However, a growing body of recent literature points to the potential of leveraging a combination of user-centered methodologies and easier user interfaces to personalize study designs and maintain high engagement through personal (remote) contact by allowing participants to build trust with researchers [92-95]. Furthermore, survey responses from 2 studies in our cohort that compared in-person and remote procedures reported high participant satisfaction with the remote interactions with investigators [43,51]. Nevertheless, a more thorough investigation on the role of in-person versus remote onboarding procedures is missing in the literature. This calls attention to a potentially essential step in remote digital health research that is understudied yet could significantly impact study participation.

### *Exploration of Hypotheses From Relationships Between Framework Criteria and Outcomes*

On average, the studies in our sample enrolled more participants than their target sample size calculations, with studies that ran for 12 weeks or less achieving higher enrollment targets than studies that ran for longer than 12 weeks. The median study completion rate was approximately 50%, which is relatively high when considering that retention rates as low as 10% are common in digital health research [7,28,96].

A description of our study sample provides indications of a possible higher study completion for studies that provide incentives or nudges. This was observed with interventional studies that, despite providing fewer lower complexity tasks from participants than the observational studies in our sample, managed to achieve higher study completion. The higher study completion could be due to interventional studies providing more incentives or nudges to participants than observational studies. The provision of incentives in studies is consistent with recent approaches that use willingness to accept estimations or incentive matching with local wages to enable higher study enrollment and completion [97-101]. Furthermore, the choice of offering nudges to study participants is supported by recent adaptations to the nudge theory, which claims that providing motivational elements in digital health research may affect decision making in study participants [102]. In recent studies, different forms of nudges, or motivational strategies, such as gamification, personalization of the digital solution, and peer support (eg, through citizen science methodologies [2]), have been reported as successful in maintaining high user retention [103-105]. The creation of online communities and support groups was also mentioned in 9 of our studies as a key contributor for participant retention [49,51,60,62,65,70,73,74,81].

Enrollment for interventional studies was higher than for observational studies. This was observed despite wider reach study recruitment and easier enrollment procedures efforts through digital recruitment channels and mobile technologies observed in most observational studies from our sample [49,54,61,75,76]. Observational studies also achieved lower completion outcomes despite providing tasks that reduce participant burden, which points to a possible lack of appropriate nudges to avoid participants dropping out after enrollment. It was also observed that studies focused on intrinsically motivated participants enrolled participants for longer studies and required them to complete more tasks, while providing less incentives or nudges than in studies focused on extrinsically motivated participants. This may have contributed to selection of specific participant profiles and higher loss to follow up, as evidenced by lower completion of studies with intrinsically motivated participants. Conversely, the studies with intrinsically motivated participants that achieved study completion values higher than the median of this group of studies predominantly focused on participants at risk or with chronic conditions in an observational study format [54,67,69,71,75]. This observation aligns with recent literature on the prominent role of disease status in enabling higher study retention in digital health studies through intrinsic motivations [19,106,107].

Our analyses yielded no statistical evidence to support our hypothesis of an interaction between factors that affect participation in remote digital health studies and the study outcomes. However, the lack of statistical support for the other criteria does not invalidate our hypotheses. First, only few remote digital health studies were identified that provided detailed information on approaches used to enhance study enrollment and retention in their methodologies. This warrants more systematic reporting of remote digital health study operations. Second, publication bias may have influenced our statistical calculations. For example, studies that failed to reach the target sample size may not have gotten published in peer-reviewed journals. This may have biased our correlations toward the null hypothesis. Third, there were observed preferences for the target participant profiles (eg, levels of digital literacy) and the study procedure that likely had an impact on the studies' outcomes. To enable future assessments that evaluate whether specific participation factors, such as mental tasks that impact cognitive burden, have an impact on study outcomes, it is recommended to report these factors individually in remote digital health studies. The availability of these data could enable more comprehensive reviews that can thoroughly investigate these interactions through statistical analyses, as well as more in-depth explorations of participation enablers and inhibitors through digital survey studies.

### *Inconsistencies of Study Completion Measures*

Heterogeneity in researchers' choices to measure study completion was observed in our sample. Although all the studies we assessed outlined the expected tasks for the study participants to complete in their methodologies (or in referenced larger studies or protocols [55,108-115]), only 3 measured study completion based on the achievement of all tasks defined in the study [64,77,81]. Interestingly, 9 studies measured study completion based on researcher-defined criteria for task completion (eg, completion of 1 task within a 30-day period) specific to their study design [49,53-55,60,61,73,75,80]. Given the novelty of digital health research, the choice of different study completion metrics by researchers is not surprising. However, this poses difficulties in allowing for direct comparisons between studies, as the former approach can be regarded as more conservative, while the latter more lenient. It is, therefore, important that study completion metrics are studied in more detail to determine whether a one-size-fits-all approach should be taken for all studies, or if bespoke approaches to measure completion are more suitable.

### Future Directions for Remote Digital Health Study Planning

Although our statistical analyses were inconclusive, we conducted an initial exploratory assessment of the study's qualitative data. We adapted the conceptual framework based on best practices found from an assessment of our study sample's descriptive results that can inform future planning of remote digital health studies:

- Adapt incentives and nudges provided to participants based on their motivation profile: offer different incentives or nudges at each key step of the study procedure. Monetary incentives may contribute to higher study enrollment [48,53,55-57,70,78], after sample validation [86,87], whereas nudges in the form of assistance during onboarding [48,55] and the provision of reminders [50,53,55,60,64,73] or a participant community (eg, through citizen science [2]) could contribute to higher retention. As technology replaces in-person interactions, the procedures set in place should be user-friendly [50,53,55-57,65,66] and enable participants to build personal relationships, with either study participants or study investigators [116]. An adequate assessment of participant profiles and their motivations to join the study can help adapt the provision of incentives or nudges. It is recommended that study investigators do not underestimate the requirement for additional incentives or nudges with intrinsically motivated participants in their study planning.
- Reduce and simplify the study's tasks: reducing the number of tasks required from participants in combination with providing incentives may achieve higher enrollment and completion in studies with intrinsically motivated participants, especially if they align with participants' daily lives [53-55,60,78,79]. The use of digital tools with simpler user interfaces and passive monitoring can also facilitate task completion by reducing participants' cognitive burden [49,54,60,73,78,79]. The implementation of onboarding procedures may also increase participants' trust, understanding of, and confidence in the study [43,51]. Based on our preliminary assessment of the upper quantile of our results, the required number of physical tasks in studies with intrinsically motivated participants should ideally not exceed 36 steps per month.
- Broaden the scientific requirements: adjust the design of the investigation and target sample size by simplifying the study's research question. Broader research questions that affect larger population groups may help with achieving a representative study sample, which can be generalized to wider populations [49,54,61,67,75]. Study designs that consider participant motivations, the provision of incentives or nudges, and task complexities can contribute to higher study enrollment and completion outcomes.

We also suggest more systematic reporting on these criteria for study enrollment and completion to facilitate further quantitative assessments and knowledge exchange. This is particularly relevant because, compared with traditional health studies, remote digital health studies bring upon additional, less well-studied challenges. To facilitate the development of remote digital health study methodologies, revised study planning approaches, as voiced previously by others [16], are necessary. This is also of relevance for hybrid or fully decentralized trials, where their reliance on remote procedures also affects participation [117,118]. There is also a need for systematic reporting of additional procedural aspects of study execution with remote digital health studies. Specifically, different reporting requirements are encouraged, such as adapted sample size calculations, approaches to remote recruitment, and onboarding of participants in the “Methods” sections of studies, as well as detailed participant descriptions and aligned reporting of study completion measures in the “Results” sections of studies. The formulation of reporting guidelines, similar to STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) [119], mERA (mobile health [mHealth] evidence reporting and assessment) guidelines [120], and CONSORT-EHEALTH standards [27], as well as further refining our conceptual framework with more evidence as it is made available, could be a first step in this direction.

## Strengths and Limitations

To our knowledge, our study is the first in the field that provides a comprehensive overview of the facilitators and barriers to participation in remote digital health studies. Our study is also the first to provide evidence-based guidelines to inform future remote digital health study planning. However, our study has limitations. First, there is no unified definition of a remote digital health study. As such, a broader definition could have yielded more studies of interest from our search strategy. Second, the analysis is reliant on inconsistent criteria for the outcomes of interest defined by the investigators of our study sample, such as approaches to calculate study completion. In this study, we grouped these criteria into 3 categories of outcome measures, however, a comparison of study outcomes with full accuracy was not possible. Third, the studies are classified based on criteria and assumptions defined by the investigators (eg, definition of task complexity), which we attempted to keep as broad as possible based on available literature in the field. Nevertheless, other variables as well as measures for the classifications and definitions could be possible. Fourth, the analyses were conducted based on a limited set of hypotheses defined in this study; more hypotheses could also be possible. Lastly, our study’s dual review approach, based on random screenings, instead of a complete dual screening, may have led to missing studies in our final study sample.

## Conclusion

In this study, we describe a conceptual framework to introduce criteria that affect remote digital health study participation from a person-centered lens. We apply this framework to remote digital health studies to explore hypotheses on the factors that affect participation outcomes. The compiled data from our scoping review reveal that targeting specific participant profiles, the provision of incentives and nudges, or the reduction of study complexity at any stage of the study may improve study outcomes. Future remote digital health study planning requires a focus on participant requirements, as well as broadening scientific requirements to increase participation in studies. Our proposed framework provides an initial structure to facilitate remote digital health study planning, but we highlight the need for systematic reporting guidelines to enable further assessments and knowledge exchange.

## Authors' Contributions

PD wrote the initial plan for the review, conducted the literature search, carried out the analysis, wrote the first draft of the paper, and wrote and revised the final manuscript. VN assisted with the initial plan for the review, conducted the literature search, helped interpret the findings, and revised and approved the final manuscript. VvW conceptualized the study, helped interpret the findings, and revised and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Methods and additional descriptive results.
[DOCX File , 27 KB - jmir_v24i9e39910_app1.docx ]

Multimedia Appendix 2
PRISMA-ScR checklist. PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-analyses Extension for Scoping Reviews.
[PDF File (Adobe PDF File), 105 KB - jmir_v24i9e39910_app2.pdf ]

Multimedia Appendix 3
Search strategy. Also see Multimedia Appendix 2.
[DOCX File , 23 KB - jmir_v24i9e39910_app3.docx ]

Multimedia Appendix 4
General characteristics of included studies.
[DOCX File , 41 KB - jmir_v24i9e39910_app4.docx ]

---

Multimedia Appendix 5
Requirements and outcomes of included studies.
[DOCX File , 30 KB - jmir_v24i9e39910_app5.docx ]

---

Multimedia Appendix 6
Incentives or nudges and task complexity of included studies.
[DOCX File , 31 KB - jmir_v24i9e39910_app6.docx ]

---

Multimedia Appendix 7
Associations between framework criteria and outcomes.
[DOCX File , 26 KB - jmir_v24i9e39910_app7.docx ]

---

## References

1. Global Smartphones Sales May Have Peaked Internet. International Monetary Fund. URL: https://www.imf.org/en/Publications/WP/Issues/2020/05/29/Global-Smartphones-Sales-May-Have-Peaked-49361 [accessed 2021-05-04]
2. Puhan MA, Steinemann N, Kamm CP, Müller S, Kuhle J, Kurmann R, Swiss Multiple Sclerosis Registry Smsr. A digitally facilitated citizen-science driven approach accelerates participant recruitment and increases study population diversity. Swiss Med Wkly 2018 May 16;148:w14623 [FREE Full text] [doi: 10.4414/smw.2018.14623] [Medline: 29767828]
3. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
4. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020;3:101 [FREE Full text] [doi: 10.1038/s41746-020-0302-y] [Medline: 32821856]
5. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917 [FREE Full text] [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
6. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
7. Crouthamel M, Quattrocchi E, Watts S, Wang S, Berry P, Garcia-Gancedo L, et al. Using a ResearchKit Smartphone App to Collect Rheumatoid Arthritis Symptoms From Real-World Participants: Feasibility Study. JMIR Mhealth Uhealth 2018 Sep 13;6(9):e177 [FREE Full text] [doi: 10.2196/mhealth.9656] [Medline: 30213779]
8. Nouri SS, Adler-Milstein J, Thao C, Acharya P, Barr-Walker J, Sarkar U, et al. Patient characteristics associated with objective measures of digital health tool use in the United States: A literature review. J Am Med Inform Assoc 2020 May 01;27(5):834-841 [FREE Full text] [doi: 10.1093/jamia/ocaa024] [Medline: 32364238]
9. Langford A, Orellana K, Kalinowski J, Aird C, Buderer N. Use of Tablets and Smartphones to Support Medical Decision Making in US Adults: Cross-Sectional Study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19531 [FREE Full text] [doi: 10.2196/19531] [Medline: 32784181]
10. Langford AT, Solid CA, Scott E, Lad M, Maayan E, Williams SK, et al. Mobile Phone Ownership, Health Apps, and Tablet Use in US Adults With a Self-Reported History of Hypertension: Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Jan 14;7(1):e12228 [FREE Full text] [doi: 10.2196/12228] [Medline: 31344667]
11. Sezgin E, Özkan-Yildirim S, Yildirim S. Investigation of physicians' awareness and use of mHealth apps: A mixed method study. Health Policy and Technology 2017 Sep;6(3):251-267. [doi: 10.1016/j.hlpt.2017.07.007]
12. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for Delivering Digital Health at Scale: Lessons From a Longitudinal Qualitative Evaluation of a National Digital Health Innovation Program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
13. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010 Dec 14;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
14. Pedersen DH, Mansourvar M, Sortsø C, Schmidt T. Predicting Dropouts From an Electronic Health Platform for Lifestyle Interventions: Analysis of Methods and Predictors. J Med Internet Res 2019 Sep 04;21(9):e13617 [FREE Full text] [doi: 10.2196/13617] [Medline: 31486409]
15. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A Self-Regulation-Based eHealth Intervention to Promote a Healthy Lifestyle: Investigating User and Website Characteristics Related to Attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]

16. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
17. Global strategy on digital health 2020-2025. In: World Health Organization. Geneva: World Health Organization; 2021.
18. Chu L, Shah A, Rouholiman D, Riggare S, Gamble J. Patient-Centric Strategies in Digital Health. In: Digital Health. Health Informatics. Cham: Springer International Publishing; 2018.
19. 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 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
20. Corona Immunitas. URL: https://www.corona-immunitas.ch/ [accessed 2021-06-24]
21. Steinemann N, Kuhle J, Calabrese P, Kesselring J, Disanto G, Merkler D, Swiss Multiple Sclerosis Registry. The Swiss Multiple Sclerosis Registry (SMSR): study protocol of a participatory, nationwide registry to promote epidemiological and patient-centered MS research. BMC Neurol 2018 Aug 13;18(1):111 [FREE Full text] [doi: 10.1186/s12883-018-1118-0] [Medline: 30103695]
22. Berwick DM. Disseminating innovations in health care. JAMA 2003 Apr 16;289(15):1969-1975. [doi: 10.1001/jama.289.15.1969] [Medline: 12697800]
23. Robinson J, Kocman D, Speyer O, Gerasopoulos E. Meeting volunteer expectations — a review of volunteer motivations in citizen science and best practices for their retention through implementation of functional features in CS tools. Journal of Environmental Planning and Management 2021 Feb 02;64(12):2089-2113. [doi: 10.1080/09640568.2020.1853507]
24. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
25. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
26. Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F. Theories Informing eHealth Implementation: Systematic Review and Typology Classification. J Med Internet Res 2021 May 31;23(5):e18500 [FREE Full text] [doi: 10.2196/18500] [Medline: 34057427]
27. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
28. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014 Feb 04;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
29. Milward J, Drummond C, Fincham-Campbell S, Deluca P. What makes online substance-use interventions engaging? A systematic review and narrative synthesis. Digit Health 2018 Feb;4:2055207617743354 [FREE Full text] [doi: 10.1177/2055207617743354] [Medline: 29942622]
30. de Ridder M, Kim J, Jing Y, Khadra M, Nanan R. A systematic review on incentive-driven mobile health technology: As used in diabetes management. J Telemed Telecare 2017 Jan;23(1):26-35. [doi: 10.1177/1357633X15625539] [Medline: 26888421]
31. Thaler R, Sunstein C. Nudge: Improving decisions about health, wealth, and happiness. In: Const Polit Econ. New Haven: Yale University Press; Aug 22, 2008:356-360.
32. Dai H, Saccardo S, Han MA, Roh L, Raja N, Vangala S, et al. Behavioural nudges increase COVID-19 vaccinations. Nature 2021 Sep 02;597(7876):404-409 [FREE Full text] [doi: 10.1038/s41586-021-03843-2] [Medline: 34340242]
33. Horne BD, Muhlestein JB, Lappé DL, May HT, Le VT, Bair TL, et al. Behavioral Nudges as Patient Decision Support for Medication Adherence: The ENCOURAGE Randomized Controlled Trial. Am Heart J 2022 Feb;244:125-134 [FREE Full text] [doi: 10.1016/j.ahj.2021.11.001] [Medline: 34798073]
34. Bauer B, Tucker R, Capron D. A Nudge in a New Direction: Integrating Behavioral Economic Strategies Into Suicide Prevention Work. Clinical Psychological Science 2019 May 01;7(3):612-620 [FREE Full text] [doi: 10.1177/2167702618809367]
35. Nittas V, Puhan MA, von Wyl V. Toward a Working Definition of eCohort Studies in Health Research: Narrative Literature Review. JMIR Public Health Surveill 2021 Jan 21;7(1):e24588 [FREE Full text] [doi: 10.2196/24588] [Medline: 33475521]
36. Nebeker C, Gholami M, Kareem D, Kim E. Applying a Digital Health Checklist and Readability Tools to Improve Informed Consent for Digital Health Research. Front Digit Health 2021 Jul 15;3:690901 [FREE Full text] [doi: 10.3389/fdgth.2021.690901] [Medline: 34713167]
37. Torous J, Michalak EE, O'Brien HL. Digital Health and Engagement-Looking Behind the Measures and Methods. JAMA Netw Open 2020 Jul 01;3(7):e2010918 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
38. Taylor KI, Staunton H, Lipsmeier F, Nobbs D, Lindemann M. Outcome measures based on digital health technology sensor data: data- and patient-centric approaches. NPJ Digit Med 2020 Jul 23;3(1):97 [FREE Full text] [doi: 10.1038/s41746-020-0305-8] [Medline: 32715091]

39. 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 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
40. El Benny M, Kabakian-Khasholian T, El-Jardali F, Bardus M. Application of the eHealth Literacy Model in Digital Health Interventions: Scoping Review. J Med Internet Res 2021 Jun 03;23(6):e23473 [FREE Full text] [doi: 10.2196/23473] [Medline: 34081023]
41. Rowsell A, Muller I, Murray E, Little P, Byrne CD, Ganahl K, et al. Views of People With High and Low Levels of Health Literacy About a Digital Intervention to Promote Physical Activity for Diabetes: A Qualitative Study in Five Countries. J Med Internet Res 2015 Oct 12;17(10):e230 [FREE Full text] [doi: 10.2196/jmir.4999] [Medline: 26459743]
42. Figueroa CA, Murayama H, Amorim PC, White A, Quiterio A, Luo T, et al. Applying the Digital Health Social Justice Guide. Front Digit Health 2022;4:807886 [FREE Full text] [doi: 10.3389/fdgth.2022.807886] [Medline: 35295620]
43. Hernandez-Ramos R, Aguilera A, Garcia F, Miramontes-Gomez J, Pathak LE, Figueroa CA, et al. Conducting Internet-Based Visits for Onboarding Populations With Limited Digital Literacy to an mHealth Intervention: Development of a Patient-Centered Approach. JMIR Form Res 2021 Apr 29;5(4):e25299 [FREE Full text] [doi: 10.2196/25299] [Medline: 33872184]
44. Barnett I, Torous J, Reeder HT, Baker J, Onnela J. Determining sample size and length of follow-up for smartphone-based digital phenotyping studies. J Am Med Inform Assoc 2020 Dec 09;27(12):1844-1849 [FREE Full text] [doi: 10.1093/jamia/ocaa201] [Medline: 33043370]
45. Whitley E, Ball J. Statistics review 4: sample size calculations. Crit Care 2002 Aug;6(4):335-341 [FREE Full text] [doi: 10.1186/cc1521] [Medline: 12225610]
46. 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]
47. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on Wearable Technology for Connected Health: Scoping Review of Research Trends, Advances, and Barriers. J Med Internet Res 2019 Sep 05;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
48. Damschroder LJ, Buis LR, McCant FA, Kim HM, Evans R, Oddone EZ, et al. Effect of Adding Telephone-Based Brief Coaching to an mHealth App (Stay Strong) for Promoting Physical Activity Among Veterans: Randomized Controlled Trial. J Med Internet Res 2020 Aug 04;22(8):e19216 [FREE Full text] [doi: 10.2196/19216] [Medline: 32687474]
49. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital Care for Chronic Musculoskeletal Pain: 10,000 Participant Longitudinal Cohort Study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
50. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant Recruitment and Retention in Remote eHealth Intervention Trials: Methods and Lessons Learned From a Large Randomized Controlled Trial of Two Web-Based Smoking Interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
51. Bott N, Kumar S, Krebs C, Glenn JM, Madero EN, Juusola JL. A Remote Intervention to Prevent or Delay Cognitive Impairment in Older Adults: Design, Recruitment, and Baseline Characteristics of the Virtual Cognitive Health (VC Health) Study. JMIR Res Protoc 2018 Aug 13;7(8):e11368 [FREE Full text] [doi: 10.2196/11368] [Medline: 30104186]
52. Hamilton FL, Hornby J, Sheringham J, Linke S, Ashton C, Moore K, et al. DIAMOND (DIgital Alcohol Management ON Demand): a feasibility RCT and embedded process evaluation of a digital health intervention to reduce hazardous and harmful alcohol use recruiting in hospital emergency departments and online. Pilot Feasibility Stud 2018;4:114 [FREE Full text] [doi: 10.1186/s40814-018-0303-7] [Medline: 29946479]
53. Abbate KJ, Hingle MD, Armin J, Giacobbi P, Gordon JS. Recruiting Women to a Mobile Health Smoking Cessation Trial: Low- and No-Cost Strategies. JMIR Res Protoc 2017 Nov 03;6(11):e219 [FREE Full text] [doi: 10.2196/resprot.7356] [Medline: 29101091]
54. Druce KL, McBeth J, van der Veer SN, Selby DA, Vidgen B, Georgatzis K, et al. Recruitment and Ongoing Engagement in a UK Smartphone Study Examining the Association Between Weather and Pain: Cohort Study. JMIR Mhealth Uhealth 2017 Nov 01;5(11):e168 [FREE Full text] [doi: 10.2196/mhealth.8162] [Medline: 29092810]
55. Gordon JS, Armin J, Giacobbi P, Cunningham JK, Johnson T, Abbate K, et al. Development and evaluation of the See Me Smoke-Free multi-behavioral mHealth app for women smokers. Transl Behav Med 2017 Feb 02;7(2):172-184. [doi: 10.1007/s13142-017-0463-7] [Medline: 28155107]
56. Laws RA, Litterbach EV, Denney-Wilson EA, Russell CG, Taki S, Ong K, et al. A Comparison of Recruitment Methods for an mHealth Intervention Targeting Mothers: Lessons from the Growing Healthy Program. J Med Internet Res 2016 Sep 15;18(9):e248. [doi: 10.2196/jmir.5691] [Medline: 27634633]
57. Ashford MT, Olander EK, Rowe H, Fisher JR, Ayers S. Feasibility and Acceptability of a Web-Based Treatment with Telephone Support for Postpartum Women With Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Apr 20;5(2):e19 [FREE Full text] [doi: 10.2196/mental.9106] [Medline: 29678804]

58. Fleischmann RJ, Harrer M, Zarski A, Baumeister H, Lehr D, Ebert DD. Patients' experiences in a guided Internet- and App-based stress intervention for college students: A qualitative study. Internet Interv 2018 Jun;12:130-140 [FREE Full text] [doi: 10.1016/j.invent.2017.12.001] [Medline: 30135777]
59. Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, et al. Efficacy of a Web-Based Guided Recommendation Service for a Curated List of Readily Available Mental Health and Well-Being Mobile Apps for Young People: Randomized Controlled Trial. J Med Internet Res 2017 May 12;19(5):e141. [doi: 10.2196/jmir.6775] [Medline: 28500020]
60. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User Engagement and Attrition in an App-Based Physical Activity Intervention: Secondary Analysis of a Randomized Controlled Trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
61. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
62. Chernick L. Improving Adolescent Sexual and Reproductive Health: Can Mobile Health Interventions Affect Behavior? Pediatrics 2021 Mar;147(3):e2020029801 [FREE Full text] [doi: 10.1542/peds.2020-029801] [Medline: 33568492]
63. Short CE, Rebar A, James EL, Duncan MJ, Courneya KS, Plotnikoff RC, et al. How do different delivery schedules of tailored web-based physical activity advice for breast cancer survivors influence intervention use and efficacy? J Cancer Surviv 2017 Feb;11(1):80-91 [FREE Full text] [doi: 10.1007/s11764-016-0565-0] [Medline: 27498099]
64. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to Internet-Based Mobile-Supported Stress Management: A Pooled Analysis of Individual Participant Data From Three Randomized Controlled Trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
65. Poppe L, Van der Mispel C, Crombez G, De Bourdeaudhuij I, Schroé H, Verloigne M. How Users Experience and Use an eHealth Intervention Based on Self-Regulation: Mixed-Methods Study. J Med Internet Res 2018 Oct 01;20(10):e10412 [FREE Full text] [doi: 10.2196/10412] [Medline: 30274961]
66. Zlotorzynska M, Bauermeister JA, Golinkoff JM, Lin W, Sanchez TH, Hightow-Weidman L. Online recruitment of youth for mHealth studies. Mhealth 2021;7:27 [FREE Full text] [doi: 10.21037/mhealth-20-64] [Medline: 33898596]
67. Garabedian LF, Ross-Degnan D, LeCates RF, Wharam JF. Uptake and use of a diabetes management program with a mobile glucometer. Prim Care Diabetes 2019 Dec;13(6):549-555 [FREE Full text] [doi: 10.1016/j.pcd.2019.03.010] [Medline: 31097343]
68. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb;41(1):74-86 [FREE Full text] [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]
69. Schneider RB, Omberg L, Macklin EA, Daeschler M, Bataille L, Anthwal S, Parkinson Study Group AT-HOME PD Investigators. Design of a virtual longitudinal observational study in Parkinson's disease (AT-HOME PD). Ann Clin Transl Neurol 2021 Feb;8(2):308-320 [FREE Full text] [doi: 10.1002/acn3.51236] [Medline: 33350601]
70. Kim H, Ray CD, Veluscek AM. Complementary Support from Facilitators and Peers for Promoting mHealth Engagement and Weight Loss. J Health Commun 2017 Nov;22(11):905-912 [FREE Full text] [doi: 10.1080/10810730.2017.1373876] [Medline: 29125393]
71. Baca-Motes K, Edwards AM, Waalen J, Edmonds S, Mehta RR, Ariniello L, et al. Digital recruitment and enrollment in a remote nationwide trial of screening for undiagnosed atrial fibrillation: Lessons from the randomized, controlled mSToPS trial. Contemp Clin Trials Commun 2019 Jun;14:100318 [FREE Full text] [doi: 10.1016/j.conctc.2019.100318] [Medline: 30656241]
72. Richards D, Murphy T, Viganó N, Timulak L, Doherty G, Sharry J, et al. Acceptability, satisfaction and perceived efficacy of "Space from Depression": an internet-delivered treatment for depression. Internet Interv 2016 Sep;5:12-22 [FREE Full text] [doi: 10.1016/j.invent.2016.06.007] [Medline: 30135802]
73. Schoenfelder E, Moreno M, Wilner M, Whitlock KB, Mendoza JA. Piloting a mobile health intervention to increase physical activity for adolescents with ADHD. Prev Med Rep 2017 Jun;6:210-213 [FREE Full text] [doi: 10.1016/j.pmedr.2017.03.003] [Medline: 28373931]
74. Schlosser DA, Campellone TR, Truong B, Anguera JA, Vergani S, Vinogradov S, et al. The feasibility, acceptability, and outcomes of PRIME-D: A novel mobile intervention treatment for depression. Depress Anxiety 2017 Jun;34(6):546-554 [FREE Full text] [doi: 10.1002/da.22624] [Medline: 28419621]
75. Pratap A, Grant D, Vegesna A, Tummalacherla M, Cohan S, Deshpande C, et al. Evaluating the Utility of Smartphone-Based Sensor Assessments in Persons With Multiple Sclerosis in the Real-World Using an App (elevateMS): Observational, Prospective Pilot Digital Health Study. JMIR Mhealth Uhealth 2020 Oct 27;8(10):e22108 [FREE Full text] [doi: 10.2196/22108] [Medline: 33107827]
76. Williamson GR, O'Connor A, Chamberlain C, Halpin D. mHealth resources for asthma and pregnancy care: Methodological issues and social media recruitment. A discussion paper. J Adv Nurs 2018 Oct;74(10):2442-2449 [FREE Full text] [doi: 10.1111/jan.13773] [Medline: 29943472]

77. Blake H, Suggs LS, Coman E, Aguirre L, Batt ME. Active8! Technology-Based Intervention to Promote Physical Activity in Hospital Employees. Am J Health Promot 2017 Mar;31(2):109-118 [FREE Full text] [doi: 10.4278/ajhp.140415-QUAN-143] [Medline: 26559712]
78. Keadle SK, Meuter L, Phelan S, Phillips SM. Charity-based incentives motivate young adult cancer survivors to increase physical activity: a pilot randomized clinical trial. J Behav Med 2021 Oct;44(5):682-693 [FREE Full text] [doi: 10.1007/s10865-021-00218-w] [Medline: 33825070]
79. Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G. Evaluating the Carrot Rewards App, a Population-Level Incentive-Based Intervention Promoting Step Counts Across Two Canadian Provinces: Quasi-Experimental Study. JMIR Mhealth Uhealth 2018 Sep 20;6(9):e178 [FREE Full text] [doi: 10.2196/mhealth.9912] [Medline: 30148712]
80. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using Mobile Apps to Assess and Treat Depression in Hispanic and Latino Populations: Fully Remote Randomized Clinical Trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
81. Edney S, Looyestyn J, Ryan J, Kernot J, Maher C. Posts, pics, or polls? Which post type generates the greatest engagement in a Facebook physical activity intervention? Transl Behav Med 2018 Nov 21;8(6):953-957 [FREE Full text] [doi: 10.1093/tbm/iby006] [Medline: 29660093]
82. Food and Drug Administration (FDA). Spectrum of Diseases/Conditions. FDA. 2000 Jan 28. URL: https://www.fda.gov/drugs/development-resources/spectrum-diseasesconditions [accessed 2022-03-16]
83. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
84. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout From an eHealth Intervention for Adults With Type 2 Diabetes: A Qualitative Study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
85. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth 2014;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
86. Pratt-Chapman M, Moses J, Arem H. Strategies for the Identification and Prevention of Survey Fraud: Data Analysis of a Web-Based Survey. JMIR Cancer 2021 Jul 16;7(3):e30730 [FREE Full text] [doi: 10.2196/30730] [Medline: 34269685]
87. Glazer JV, MacDonnell K, Frederick C, Ingersoll K, Ritterband LM. Liar! Liar! Identifying eligibility fraud by applicants in digital health research. Internet Interv 2021 Sep;25:100401 [FREE Full text] [doi: 10.1016/j.invent.2021.100401] [Medline: 34094883]
88. Kolovson S, Pratap A, Duffy J, Allred R, Munson SA, Areán PA. Understanding Participant Needs for Engagement and Attitudes towards Passive Sensing in Remote Digital Health Studies. Int Conf Pervasive Comput Technol Healthc 2020 May;2020:347-362 [FREE Full text] [doi: 10.1145/3421937.3422025] [Medline: 33717638]
89. Baumel A, Muench FJ. Effort-Optimized Intervention Model: Framework for Building and Analyzing Digital Interventions That Require Minimal Effort for Health-Related Gains. J Med Internet Res 2021 Mar 12;23(3):e24905 [FREE Full text] [doi: 10.2196/24905] [Medline: 33709943]
90. Druce KL, Dixon WG, McBeth J. Maximizing Engagement in Mobile Health Studies: Lessons Learned and Future Directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
91. Rawstorn JC, Gant N, Rolleston A, Whittaker R, Stewart R, Benatar J, et al. End Users Want Alternative Intervention Delivery Models: Usability and Acceptability of the REMOTE-CR Exercise-Based Cardiac Telerehabilitation Program. Arch Phys Med Rehabil 2018 Nov;99(11):2373-2377 [FREE Full text] [doi: 10.1016/j.apmr.2018.06.027] [Medline: 30076800]
92. Partridge S, Redfern J. Strategies to Engage Adolescents in Digital Health Interventions for Obesity Prevention and Management. Healthcare 2018 Jun 21;6(3):70 [FREE Full text] [doi: 10.3390/healthcare6030070] [Medline: 29933550]
93. Tebb KP, Leng Trieu S, Rico R, Renteria R, Rodriguez F, Puffer M. A Mobile Health Contraception Decision Support Intervention for Latina Adolescents: Implementation Evaluation for Use in School-Based Health Centers. JMIR Mhealth Uhealth 2019 Mar 14;7(3):e11163 [FREE Full text] [doi: 10.2196/11163] [Medline: 30869649]
94. Dulli L, Ridgeway K, Packer C, Murray KR, Mumuni T, Plourde KF, et al. A Social Media-Based Support Group for Youth Living With HIV in Nigeria (SMART Connections): Randomized Controlled Trial. J Med Internet Res 2020 Jun 02;22(6):e18343 [FREE Full text] [doi: 10.2196/18343] [Medline: 32484444]
95. Verbiest MEA, Corrigan C, Dalhousie S, Firestone R, Funaki T, Goodwin D, et al. Using codesign to develop a culturally tailored, behavior change mHealth intervention for indigenous and other priority communities: A case study in New Zealand. Transl Behav Med 2019 Jul 16;9(4):720-736 [FREE Full text] [doi: 10.1093/tbm/iby093] [Medline: 30388262]
96. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of Obtaining Measures of Lifestyle From a Smartphone App: The MyHeart Counts Cardiovascular Health Study. JAMA Cardiol 2017 Jan 01;2(1):67-76 [FREE Full text] [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
97. Huberman B, Adar E, Fine L. Valuating Privacy. IEEE Secur. Privacy Mag 2005 Sep;3(5):22-25 [FREE Full text] [doi: 10.1109/msp.2005.137]

98. Heidel A, Hagist C, Schlereth C. Pricing through health apps generated data-Digital dividend as a game changer: Discrete choice experiment. PLoS One 2021;16(7):e0254786 [FREE Full text] [doi: 10.1371/journal.pone.0254786] [Medline: 34310618]
99. Vlaev I, King D, Darzi A, Dolan P. Changing health behaviors using financial incentives: a review from behavioral economics. BMC Public Health 2019 Aug 07;19(1):1059 [FREE Full text] [doi: 10.1186/s12889-019-7407-8] [Medline: 31391010]
100. Promberger M, Dolan P, Marteau TM. "Pay them if it works": discrete choice experiments on the acceptability of financial incentives to change health related behaviour. Soc Sci Med 2012 Dec;75(12):2509-2514 [FREE Full text] [doi: 10.1016/j.socscimed.2012.09.033] [Medline: 23102753]
101. Short K, Chadwick J, Cannady T, Branam D, Wharton D, Tullier M, et al. Using financial incentives to promote physical activity in American Indian adolescents: A randomized controlled trial. PLoS ONE 2018 Jun 1;13(6):e0198390 [FREE Full text] [doi: 10.1371/journal.pone.0198390] [Medline: 29856832]
102. Voyer B. 'Nudging' behaviours in healthcare: Insights from behavioural economics. British Journal of Healthcare Management 2015 Mar 02;21(3):130-135 [FREE Full text] [doi: 10.12968/bjhc.2015.21.3.130]
103. van Roy R, Zaman B. Why Gamification Fails in Education and How to Make It Successful: Introducing Nine Gamification Heuristics Based on Self-Determination Theory. In: Serious Games and Edutainment Applications. Cham: Springer International Publishing; 2017. [doi: 10.1007/978-3-319-51645-5_22]
104. Martin A, Caon M, Adorni F, Andreoni G, Ascolese A, Atkinson S, et al. A Mobile Phone Intervention to Improve Obesity-Related Health Behaviors of Adolescents Across Europe: Iterative Co-Design and Feasibility Study. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e14118 [FREE Full text] [doi: 10.2196/14118] [Medline: 32130179]
105. 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 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
106. Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007 May 09;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]
107. 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 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
108. Denney-Wilson E, Laws R, Russell CG, Ong K, Taki S, Elliot R, et al. Preventing obesity in infants: the Growing healthy feasibility trial protocol. BMJ Open 2015 Nov 30;5(11):e009258 [FREE Full text] [doi: 10.1136/bmjopen-2015-009258] [Medline: 26621519]
109. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: Its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31 [FREE Full text] [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
110. Richards D, Timulak L, Doherty G, Sharry J, Colla A, Joyce C, et al. Internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression: protocol for a randomized controlled trial. BMC Psychiatry 2014 May 21;14:147 [FREE Full text] [doi: 10.1186/1471-244X-14-147] [Medline: 24886179]
111. Bricker JB, Mull KE, McClure JB, Watson NL, Heffner JL. Improving quit rates of web-delivered interventions for smoking cessation: full-scale randomized trial of WebQuit.org versus Smokefree.gov. Addiction 2018 May;113(5):914-923 [FREE Full text] [doi: 10.1111/add.14127] [Medline: 29235186]
112. Edney S, Plotnikoff R, Vandelanotte C, Olds T, De Bourdeaudhuij I, Ryan J, et al. "Active Team" a social and gamified app-based physical activity intervention: randomised controlled trial study protocol. BMC Public Health 2017 Nov 02;17(1):859 [FREE Full text] [doi: 10.1186/s12889-017-4882-7] [Medline: 29096614]
113. Ybarra M, Goodenow C, Rosario M, Saewyc E, Prescott T. An mHealth Intervention for Pregnancy Prevention for LGB Teens: An RCT. Pediatrics 2021 Mar;147(3):e2020013607 [FREE Full text] [doi: 10.1542/peds.2020-013607] [Medline: 33568491]
114. Looyestyn J, Kernot J, Boshoff K, Maher C. A Web-Based, Social Networking Beginners' Running Intervention for Adults Aged 18 to 50 Years Delivered via a Facebook Group: Randomized Controlled Trial. J Med Internet Res 2018 Feb 26;20(2):e67 [FREE Full text] [doi: 10.2196/jmir.7862] [Medline: 29483065]
115. Heber E, Ebert DD, Lehr D, Nobis S, Berking M, Riper H. Efficacy and cost-effectiveness of a web-based and mobile stress-management intervention for employees: design of a randomized controlled trial. BMC Public Health 2013 Jul 15;13:655 [FREE Full text] [doi: 10.1186/1471-2458-13-655] [Medline: 23855376]
116. Dalton C, Carlson S, Butler M, Cassano D, Clarke S, Fejsa J, et al. Insights From Flutracking: Thirteen Tips to Growing a Web-Based Participatory Surveillance System. JMIR Public Health Surveill 2017 Aug 17;3(3):e48 [FREE Full text] [doi: 10.2196/publichealth.7333] [Medline: 28818817]
117. De Brouwer W, Patel CJ, Manrai AK, Rodriguez-Chavez IR, Shah NR. Empowering clinical research in a decentralized world. NPJ Digit Med 2021 Jul 01;4(1):102 [FREE Full text] [doi: 10.1038/s41746-021-00473-w] [Medline: 34211085]

118. Apostolaros M, Babaian D, Corneli A, Forrest A, Hamre G, Hewett J, et al. Legal, Regulatory, and Practical Issues to Consider When Adopting Decentralized Clinical Trials: Recommendations From the Clinical Trials Transformation Initiative. Ther Innov Regul Sci 2020 Jul;54(4):779-787 [FREE Full text] [doi: 10.1007/s43441-019-00006-4] [Medline: 32557302]
119. Field N, Cohen T, Struelens MJ, Palm D, Cookson B, Glynn JR, et al. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. Lancet Infect Dis 2014 Apr;14(4):341-352 [FREE Full text] [doi: 10.1016/S1473-3099(13)70324-4] [Medline: 24631223]
120. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174 [FREE Full text] [doi: 10.1136/bmj.i1174] [Medline: 26988021]

## Abbreviations

**CONSORT-EHEALTH:** Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth
**mERA:** mobile health (mHealth) evidence reporting and assessment
**PRISMA-ScR:** Preferred Reporting Items for Systematic reviews and Meta-analyses Extension for Scoping Reviews
**STROBE:** STrengthening the Reporting of OBservational studies in Epidemiology

*Edited by T Leung; submitted 27.05.22; peer-reviewed by DS Morgan, M Tizzani, K Reuter; comments to author 21.06.22; revised version received 12.07.22; accepted 31.07.22; published 09.09.22.*

*Please cite as:*
*Daniore P, Nittas V, von Wyl V*
*Enrollment and Retention of Participants in Remote Digital Health Studies: Scoping Review and Framework Proposal*
*J Med Internet Res 2022;24(9):e39910*
*URL: https://www.jmir.org/2022/9/e39910*
*doi:10.2196/39910*
*PMID:36083626*

Review

# Evaluating the Effectiveness of Internet-Based Communication for Public Health: Systematic Review

Elisabetta Ceretti[1*], PhD; Loredana Covolo[1*], ScD; Francesca Cappellini[1], MA; Alberto Nanni[2], MD; Sara Sorosina[2], MA; Andrea Beatini[2], MA; Mirella Taranto[3], MA; Arianna Gasparini[3], MA; Paola De Castro[3], MA; Silvio Brusaferro[3], MD, PhD; Umberto Gelatti[1], MD, PhD

[1]Section of Public Health and Human Sciences, Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, University of Brescia, Brescia, Italy

[2]Post-graduate School of Public Health, University of Brescia, Brescia, Italy

[3]Italian National Institute of Health, Rome, Italy

[*]these authors contributed equally

**Corresponding Author:**
Francesca Cappellini, MA
Section of Public Health and Human Sciences, Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, University of Brescia
Viale Europa 11
Brescia, 25123
Italy
Phone: 39 3381617907
Email: francesca.cappellini@unibs.it

## Abstract

**Background:** Communicating strategically is a key issue for health organizations. Over the past decade, health care communication via social media and websites has generated a great deal of studies examining different realities of communication strategies. However, when it comes to systematic reviews, there is fragmentary evidence on this type of communication.

**Objective:** The aim of this systematic review was to summarize the evidence on web institutional health communication for public health authorities to evaluate possible aim-specific key points based on these existing studies.

**Methods:** Guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement, we conducted a comprehensive review across 2 electronic databases (PubMed and Web of Science) from January 1, 2011, to October 7, 2021, searching for studies investigating institutional health communication. In total, 2 independent researchers (AN and SS) reviewed the articles for inclusion, and the assessment of methodological quality was based on the Kmet appraisal checklist.

**Results:** A total of 78 articles were selected. Most studies (35/78, 45%) targeted health promotion and disease prevention, followed by crisis communication (24/78, 31%), general health (13/78, 17%), and misinformation correction and health promotion (6/78, 8%). Engagement and message framing were the most analyzed aspects. Few studies (14/78, 18%) focused on campaign effectiveness. Only 23% (18/78) of the studies had an experimental design. The Kmet evaluation was used to distinguish studies presenting a solid structure from lacking studies. In particular, considering the 0.75-point threshold, 36% (28/78) of the studies were excluded. Studies above this threshold were used to identify a series of aim-specific and medium-specific suggestions as the communication strategies used differed greatly.

**Conclusions:** Overall, the findings suggest that no single strategy works best in the case of web-based health care communication. The extreme variability of outcomes and the lack of a unitary measure for assessing the end points of a specific campaign or study lead us to reconsider the tools we use to evaluate the efficacy of web-based health communication.

***(J Med Internet Res 2022;24(9):e38541)*** doi:10.2196/38541

**KEYWORDS**

internet-based communication; websites; social media; public health; efficacy; systematic review; communication; internet-based; health information; exchange; health care; web-based; campaigns

# Introduction

## Background

Communicating strategically requires a clearly defined strategy with specific goals established in advance. The core agenda of strategic communication is the analysis and explanation of intentional and purposeful communicative relationships between organizations and the public [1]. That being said, it is important to point out that organizations make strategic decisions about the level and nature of resources they devote to such efforts, modulating their tone and tenor of communication depending on the audience they want to appeal to [2]. In the field of institutional health care communication, this theme is especially relevant, as illustrated by the *WHO fact sheet on the Strategic Communications Framework for Effective Communications* [3]. In this document, the World Health Organization wanted to establish a framework to describe a strategic approach for effectively communicating health care–related information, advice, and guidance across a broad range of health issues. This resulted in the identification of 6 key principles: accessibility, actionability, credibility and trustworthiness, relevancy, timeliness, and understandability. However, we can observe that these guidelines are not specific enough and, on the contrary, appear to be too broad. Looking at the existing literature, it is also possible to observe a lack of specific evidence regarding the effectiveness of those studies on institutional health care communication. Therefore, it is important to be able to effectively communicate with the public at large. This would allow public health officials to minimize damage and possibly prevent widespread illness and diseases. Providing accurate and verifiable information is also paramount to keep the public informed and allow them to take the appropriate action. One of the main aims of this systematic review was to analyze a corpus of studies on institutional health care communication to see whether it is possible to extrapolate aim-specific key points based on these existing studies.

As of January 2021, there were 4.66 billion active internet users on the web; 59.5% of the entire population [4]. With the dramatic increase in internet access, there has been a parallel increase in the use of the internet as a platform for the delivery of public health interventions across a wide range of conditions and population segments [5]. Over the past decade, health care communication via social media and websites has generated a great deal of studies examining different realities of communication strategies [6-8]. However, this vast diffusion of internet health care communication is a double-edged sword, as demonstrated by the *infodemic* [9] occurring during the ongoing COVID-19 pandemic. In this context, along with the diffusion of trustworthy information and guidelines from governments and health care organizations, a massive wave of false information has also spread. Although misinformation has spread throughout history, social media and technological advances in communication have amplified its impact, making it difficult for information from official sources to spread effectively without being drowned by this false information [10]. Thus, the absence of specific guidelines to effectively communicate via social media or websites has posed a problem that is yet to be addressed properly, as public health institutions have struggled to find their footing in this area, as well as a unified communication strategy for the diffusion of official messages [11].

The current evidence on internet-based health care communication appears to be rather fragmentary and localized according to topic- and platform-specific criteria. A number of other systematic reviews were published over the past 10 years [12-18]. In particular, the systematic review by Moorhead et al [18] claims that there is a lack of communication about the uses, beliefs, and limitations of social media for health communication. In total, 2 other systematic reviews [14,17] deal with providing evidence of effectiveness for studies on web-based communication, concluding that effectiveness was only sparsely reported and reach was only being assessed among those involved in the research process. Going into even more specific accounts as related to web-based health communication, the studies by Alamoodi et al [12], Kim [15], and Lehto and Oinas-Kukkonen [16] deal with the public's perception of this type of communication in 3 different instances: trust in websites, persuasive features of web-based interventions, and application of sentiment analysis. Even in this case, the conclusions leave a substantial gap to be filled with future research. One last study [13] focuses on a completely different aspect of social media communication by basing its text collection on studies on specific social media platforms rather than on specific interventions made on the web at large. More specifically, picture-based social media such as Instagram, Pinterest, Tumblr, and Flickr are the platforms taken into account. In this case, the focus is on images used as vehicles for health care communication. However, most of these studies appear to be observational, and only few provide more specific intervention tools.

## Objectives

The objective of this systematic review was to form a more comprehensive and extensive account on the matter of web-based health communication (especially making reference to national health care institutions and nongovernmental organizations) than the aforementioned studies through a comprehensive bibliographic search of articles dealing with this topic over multiple platforms. In addition to identifying the most relevant articles on this matter, this review tried to define a series of key points as comprehensively as possible that can be applied to health campaigns spread through websites or different social media by health organizations.

# Methods

This systematic review was carried out according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [19].

## Information Sources

The literature search covered the period from January 1, 2011, to October 7, 2021—as web-based communication has undergone a rapid and drastic change over the past decade and research published before this date can appear to be rather obsolete and misleading for the scope of this study—and was carried out using electronic databases. The research process was separated into 2 parts: (1) research via electronic databases (PubMed and Web of Science) and (2) research through analysis of relevant systematic reviews (bibliographies were analyzed, and suitable articles were assessed for eligibility).

### Search Strategies and Study Selection

A bibliographic search was conducted on PubMed and Web of Science using the following search string: (*Social Media* OR *Twitter* OR *Facebook* OR *Instagram* OR *Website*) AND (*communication strategy**) AND (*health* OR *public health* OR *organization** OR *agenc** OR *risk*) NOT (*hospital** OR *practitioner**).

Duplicates were identified via Zotero (Corporation for Digital Scholarship) [20,21] and eliminated.

Search results were initially evaluated based on the title and abstract by 2 independent reviewers (AN and SS), which resulted in the exclusion of all clearly irrelevant articles. In case of disagreement between the 2 parties, a third member of the team (FC) was included to resolve all conflicts.

All studies identified in this preliminary evaluation phase were considered eligible for assessment based on the exclusion and inclusion criteria stated in the following section.

### Inclusion and Exclusion Criteria

We included articles according to the following criteria: peer-reviewed or book section; published between January 1, 2011 and October 7, 2021; and written in English.

As for the research topic, we included research papers focused on social media– or website-based institutional communication strategies for health care promotion and health care promotion campaigns organized by public authorities or health care–related nongovernmental organizations spread via social media or websites and that illustrated their communication strategies.

We excluded all publications related to communication strategies applied to physician-patient communication, telemedicine, and hospital portals addressing patients; articles related to marketing communication and private institutions were also left out. The exclusion criteria also comprised qualitative studies and preliminary and exploratory articles.

### Quality Assessment

The methodological quality of each study was assessed by 2 of the authors (FC and AB) using the Kmet tool for evaluating quantitative and qualitative research [22]. A score between 0 and 1 was assigned to each paper based on a series of questions related to the type of study. Examples of items include the following: description of the research objective, appropriateness of the study design, description of participant characteristics, blinding, sample size, analytic methods, estimates of variance, control of confounding factors, and reporting of results and conclusions. A score of >0.75% was considered good quality, 0.55% to 0.75% was considered adequate quality, and <0.55% was considered poor quality. Any disagreements were resolved through discussion among the authors until a consensus was reached. Interrater reliability for the Kmet ratings was established based on κ calculations.

To further analyze the difference in the distribution of studies according to their quality, chi-square or Fischer exact tests were carried out analyzing the differences between the number of studies above and below the 0.75-point threshold.

### Data Collection and Analysis

We categorized the studies into 4 groups according to the topic addressed: crisis communication, health promotion and disease prevention, general health, and misinformation correction and health literacy. For public health emergencies, risk communication includes a range of communication capacities with the aim of encouraging positive decision-making, positive behavior change, and the maintenance of trust. This definition seems to be applicable to both the crisis communication and health promotion and disease prevention categories [23]. However, there is an important difference in the aims of these 2 types of communication: in the case of health promotion and disease prevention, health messaging advocates for an ongoing behavior change (ie, a behavior that requires an individual to keep up with a habitual activity); differently, in the case of crisis communication, the behavior change that is promoted is episodic and valid only in the case of a specific emergency [24]. Finally, those studies not dealing with any of the aforementioned categories were classified under general health. This was the case for studies analyzing the impact of a certain communication theory on communication or studies that globally analyzed a certain communication medium.

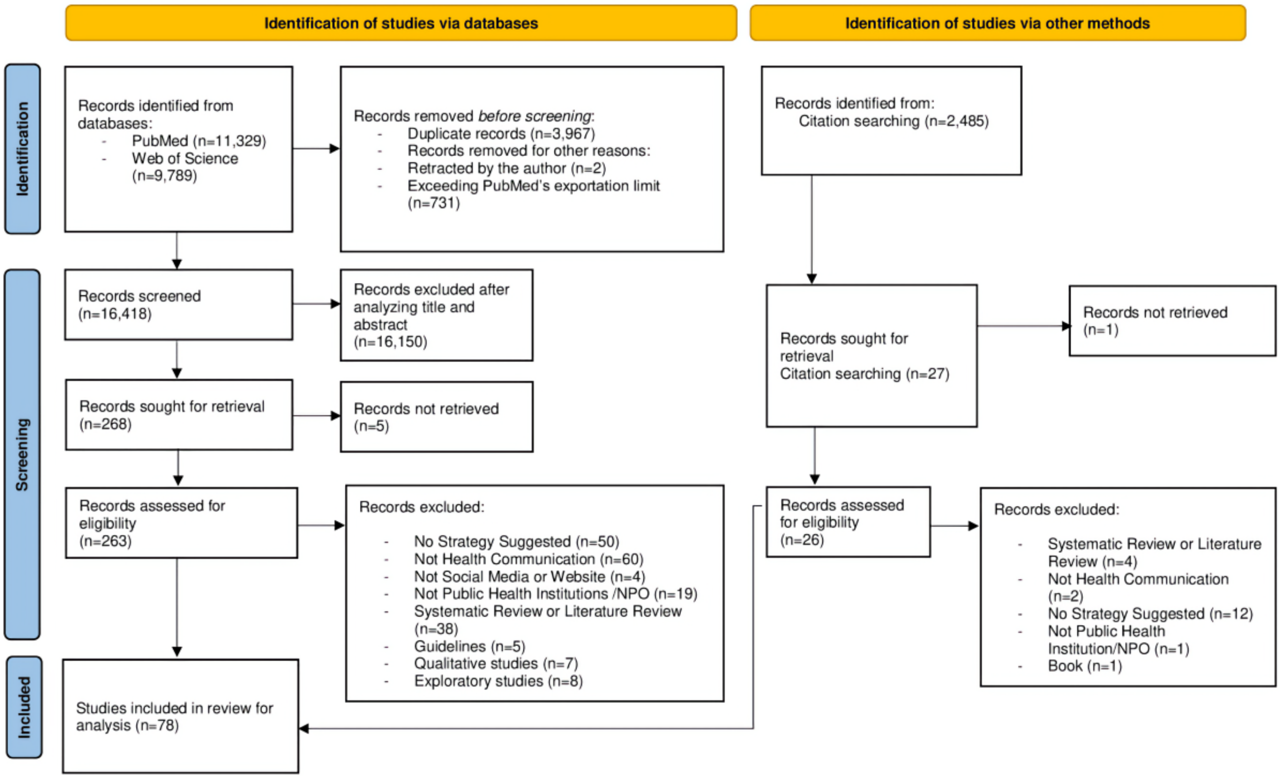
We further categorized articles according to their primary evaluation aspects. These are *engagement*, *message framing*, and *campaign effectiveness*. First, *engagement* is defined as a psychological and behavioral attribute of connection, interaction, participation, and involvement designed to elicit an outcome at the individual or social level [25]. In particular, in the case of social media, it is closely related to the concept of interaction with posts, where engagement is measured as the sum of the number of likes, comments, and shares [26]. Second, *campaign effectiveness* is closely related to the change in one’s attitudes and behaviors regarding a certain issue [27]. Finally, *message framing* constitutes the way in which a certain message is expressed and carried out (eg, gain- or loss-framed messages), and its content and connotative structure can prove effective in motivating individuals to engage in health-related behaviors [28].

Regarding study design, we categorized as *experimental* those studies where a specific intervention was recorded. More specifically, this can mean subjecting a group of individuals to different iterations of a post to see how its framing affects them. In the *observational* category, we included cross-sectional studies aimed at analyzing how a population sample reacted to a specific intervention (eg, the implementation of a certain campaign). Finally, *content analysis* refers to the analysis of a specific collection of posts with regard to their characteristics and the engagement generated.

To further analyze the effects of the threshold applied to the studies in this systematic review, the Fischer exact test was carried out analyzing the differences between the number of studies above and below the 0.75-point cutoff for all communication media.

Reported in Figure 1 is the PRISMA flow diagram for this specific systematic review.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for this systematic review.



# Results

## Overview

Of the 288 relevant articles selected, 78 (27.1%) met the inclusion criteria and were considered for this systematic review. These articles were divided into 4 categories according to the primary evaluated aspect of the study: (1) studies on crisis communication, (2) studies on health promotion and disease prevention, (3) studies on general health, and (4) studies on health literacy and misinformation correction. In particular, the latter category comprised studies on topics that appeared only a few times throughout the corpus, making it difficult to group them by themselves. Overall, the Kmet score of the evidence reviewed ranged from 0.40 to 0.93, with an average score of 0.75 (SD 0.10) and a correlation coefficient of 0.80 between the 2 reviewers.

## Studies on Crisis Communication

### *Overview*

We selected 24 studies dealing with health care institution interventions on crisis communication (Table 1), of which 12 (50%) were carried out in the United States, 3 (13%) were carried out in China, and 3 (13%) were carried out in Canada. Engagement was the most represented primary evaluated aspect (17/24, 71%), followed by message framing (4/24, 17%) and campaign effectiveness (3/24, 13%). In this category, the Kmet evaluation score resulted in an average of 0.73 points, with a correlation coefficient of 0.85.

Overall, in this group, studies focused for the most part on the analysis of collections of posts and inquiries (21/24, 88%), whereas the rest (3/24, 13%) focused on the analysis of people as participants. Consequently, the design of these studies included a high percentage of content analyses (14/24, 58%), observational studies (5/24, 21%), and network analyses (4/24, 17%). As for the media channels analyzed in this group, 83% (20/24) of the studies focused on only 1 communication medium, whereas 17% (4/24) dealt with multiple media. Finally, half of the studies (12/24, 50%) referred to a specific communication theory.

**Table 1.** List of studies on crisis communication (see Multimedia Appendix 1 [24-101] for more details; N=24).

| Primary evaluated aspect and communication medium | Reference | Studies, n (%) |
|---|---|---|
| **Engagement** | | |
| Facebook | • Dimanlig-Cruz et al [31]<br>• Lwin et al [37] | 2 (8) |
| Twitter | • Kim et al [25]<br>• Dimanlig-Cruz et al [31]<br>• Hagen et al [32]<br>• Lauran et al [33]<br>• Slavik et al [34]<br>• McInnes and Hornmoen [35]<br>• Vos et al [36]<br>• Sutton et al [38]<br>• Young et al [40]<br>• Guidry et al [42]<br>• Renshaw et al [43]<br>• Vos et al [44] | 12 (50) |
| Instagram | • Dimanlig-Cruz et al [31]<br>• Guidry et al [42] | 2 (8) |
| Other social media (Sina Weibo, TikTok, and YouTube) | • Chen et al [26,30]<br>• Dimanlig-Cruz et al [31] | 3 (13) |
| **Message framing** | | |
| Facebook | • Jang and Baek [47] | 1 (4) |
| Twitter | • Sutton et al [38]<br>• Pascual-Ferrá et al [45] | 2 (8) |
| Website | • Ort and Fahr [46] | 1 (4) |
| Other social media (Kakao Talk) | • Jang and Baek [47] | 1 (4) |
| **Campaign effectiveness** | | |
| Facebook | • MacKay et al [29]<br>• Duong et al [49] | 2 (8) |
| Website | • Harris-Sagaribay et al [50] | 1 (4) |
| Other social media (YouTube and Zalo) | • Duong et al [49] | 1 (4) |

### *Engagement*

In 71% (17/24) of the studies, the primary aim was to assess the success of engagement techniques in web-based communication both on websites and social media.

First, what emerged in the study by MacKay et al [29] was that public health agencies and news media should use guiding principles consistently to increase positive sentiment and build trust among followers.

Another study by Alamoodi et al [12] was focused on TikTok with the aim of determining the factors and influencing mechanisms related to citizen engagement with the TikTok account of the National Health Commission of China during the COVID-19 pandemic. The result of this was that shorter videos are preferred to longer ones, and a positive emotion is better suited than a negative one. Similarly, a study carried out in China [30], this time on the platform Sina Weibo, concluded that posts displaying positive emotions can include more videos or pictures, whereas plain text is more suitable for posts with negative emotions.

The studies by Dimanlig-Cruz et al [31], Hagen et al [32], Lauran et al [33], and Slavik et al [34] dealt with targeting specific population groups. The first and most generalizable study is the one by Lauran et al [33], who stated that deciding on 1 actor and 1 (homogeneous) stakeholder group is not the right strategy. What is advisable is to take the perspectives of the multiple stakeholders into account (and find opinion leaders within those groups) when deciding on the communication strategies to use and to refrain from introducing a new, unrelated issue into the discussion before the original issue is handled. Similarly, according to Hagen et al [32], public health organizations can benefit from understanding the types of content that are transmitted through specific social media platforms and identifying key participants who are authoritative, popular, and connected with disparate communities to efficiently communicate with the public. As for the study by Dimanlig-Cruz et al [31], given the high number of youths on Instagram and YouTube, public health officials may want to consider targeting youths on these sites; similarly, Slavik et al [34] tried to assess tweeting practices during public health crises to improve risk communication and maximize engagement. What emerged was the need for public health agencies to monitor Twitter analytics to understand their audience and leverage whatever Twitter engagement strategies help maximize the shares of their communications.

Creating a community was also a very important point in these studies as coordinating communication efforts by frequently interacting with other organizations to boost one's network position can facilitate further communication efforts [25]. In particular, what emerged is that organizations should consider retweeting content from health information sources with a high number of Twitter followers if they want to build up their own follower base and that health agencies should coordinate their communication efforts by frequently interacting with each other. This will boost their network position and facilitate further communication efforts. Another key strategy for public health agencies might be to develop a community of trusted users with their own significant base of followers who will pass on tweets from health authorities [35]. In the event of an outbreak, prompt responses from the authorities can be vital in crisis management, as explained by Vos et al [36], who stated that public health officials may want to emphasize the severity of an emerging infectious disease. Efficacy information is an important message element in encouraging an effective response. Precise guidelines have also been proposed in the event of a specific outbreak (COVID-19) or with regard to specific communication channels (Twitter). In the case of COVID-19, Lwin et al [37] focused on the dissemination of posts regarding the COVID-19 pandemic, and their findings showed that the public liked and shared the most in the preoutbreak phase and engaged with posts much less during the outbreak, as well as the fact that the public liked the most the posts that encouraged self-efficacy. Furthermore, in an uncertain environment, public agencies can reach the public—and increase message sharing—with a wide range of practical information regarding the health impacts of COVID-19, protective action measures, and the progress of the pandemic itself. At the same time, some tactics useful in other disasters (such as sentence styles that use exclamatory and interrogative punctuation) were counterproductive during the COVID-19 pandemic [38]. As for studies on Twitter, according to Tang et al [39], the main takeaway was that public health agencies should continue to use Twitter to disseminate information, promote action, and build communities, especially by targeting specific population groups. Similarly, Young et al [40] focused on chats, concluding that this means of communication was effective at answering questions about disease, creating a forum for targeted criticism, and promoting conversation among participants. Government accounts could also take full advantage of social media functions, especially mentions, hashtags, and the number of original posts, and add pictures and text length appropriately to increase interactions with the public and improve the level of engagement [41].

As for the strategies that proposed taking more technical aspects into account, the studies by Guidry et al [42], Renshaw et al [43], and Vos et al [44] offered interesting insights. First, according to Renshaw et al [43], focusing on useful content rather than gimmicks to go viral would be helpful in the long run. Having meaningful content such as relevant images embedded in posts might be crucial for success and, according to Guidry et al [42] and Vos et al [44], organizations should create messages that illustrate information visually and try to include threat and efficacy information in messages. They should also engage social media audiences before public health crises emerge.

### *Message Framing*

This group comprises 17% (4/24) of the studies, all aimed at assessing the way in which a certain message is framed to make communication as effective as possible. To begin with, Pascual-Ferrá et al [45] concluded that the integration of social network analysis is recommended as a best practice in crisis communication on social media. Ort and Fahr [46] conducted a study focused on the interaction between perception of threat and self-efficacy in a crisis situation. Even in this case, health messages promoting people's self-efficacy perceptions may be preferable to threatening messages. Another study aimed at a specific part of the population—public health officials—carried out by Jang and Baek [47] in South Korea concluded that lower perceived credibility of information from public health officials was associated with a greater tendency to use web-based news, interpersonal networks, and social media. The last study [48] focused on how message construction, style, content, and the textual content of tweets and embedded images affected message retransmission over the course of the pandemic. In particular, the focus was on tweet effectiveness, this time studying hashtags and concluding that public health–oriented hashtag campaigns may help engage individuals to help them feel part of a larger collective body and participate locally by contributing information about their local context.

### *Campaign Effectiveness*

The last group in this category comprised 8% (2/24) of the studies, aimed at assessing the success of certain campaigns. First, Duong et al [49] investigated the content and format of physical distancing messages directed at Vietnamese youths during the COVID-19 pandemic, concluding that perceived norms and self-efficacy did not fully account for the association between interpersonal communication and behavioral intentions. Second, Harris-Sagaribay et al [50] summarized the lessons learned through an observational retrospective study when it comes to improving information dissemination during a health care crisis. Other than content, the effectiveness of website-based communication was measured through ease of navigation and trust in the information provided by the website.

## Studies on Disease Prevention and Health Promotion

### *Overview*

This section comprises 45% (35/78) of the articles, all concerning studies that deal with the themes of health promotion and disease prevention (Table 2). Of these 35 studies, 19 (54%) were carried out in the United States, 5 (14%) were carried out in Australia, and 4 (11%) were carried out in the United Kingdom. Most studies had to do with engagement (19/35, 54%), 23% (8/35) had to do with campaign effectiveness strategies and, finally, 23% (8/35) had to do with message framing. The Kmet evaluation resulted in an average of 0.77 points, with a correlation coefficient of 0.78.

Studies on these 2 topics were mostly based on the analysis of collections of posts (21/35, 60%) rather than being studies on groups of people (14/35, 40%). The study designs were also diversified in this group, with the most represented being content analyses (16/35, 46%) and observational studies (9/35, 26%). In this group, most studies (27/35, 77%) were conducted on only 1 communication medium, whereas the remaining 23% (8/35) dealt with multiple platforms. As for communication theories, 71% (25/35) of the studies referred to one or more specific communication theories, and 29% (10/35) did not.

**Table 2.** List of studies on disease prevention and health promotion (see Multimedia Appendix 1 for more details; N=35).

| Primary evaluated aspect and communication medium | Reference | Studies, n (%) |
|---|---|---|
| **Engagement** | | |
| Facebook | • Alonso-Cañadas et al [51]<br>• Loft et al [52]<br>• Syred et al [53]<br>• Zhang and Zhou [54]<br>• Kite et al [56]<br>• Lister et al [57]<br>• Parackal et al [58]<br>• Reuter et al [60]<br>• Barklamb et al [61]<br>• Klassen et al [63]<br>• Rus and Cameron [66]<br>• Strekalova and Krieger [67]<br>• Theiss et al [68] | 13 (37) |
| Twitter | • Zhang et al [55]<br>• Lister et al [57]<br>• Rabarison et al [59]<br>• Reuter et al [60]<br>• Kim et al [62]<br>• Guidry et al [64]<br>• Chung [65] | 7 (20) |
| Instagram | • Reuter et al [60]<br>• Barklamb et al [61]<br>• Klassen et al [63]<br>• Alkazemi et al [69] | 4 (11) |
| Website | • Lister et al [57] | 1 (3) |
| Other social media (anonymous discussion platform) | • Zhang et al [55] | 1 (3) |
| **Message framing** | | |
| Facebook | • Dockter et al [24]<br>• Borah and Xiao [28]<br>• Yoo et al [70]<br>• Parackal et al [72] | 4 (11) |
| Twitter | • Yoo et al [70]<br>• Cho et al [71]<br>• Chung and Lim [74] | 2 (6) |
| Instagram | • Yoo et al [70]<br>• Nobles et al [73] | 2 (6) |
| Website | • Chung and Lim [74]<br>• Whitten et al [75] | 3 (9) |
| Other social media (YouTube, Flickr, Kakao Story, and Naver Band) | • Yoo et al [70] | 1 (3) |
| **Campaign effectiveness** | | |
| Facebook | • Potente et al [27] | 1 (3) |
| Twitter | • Allen et al [77]<br>• Yoo et al [79] | 2 (6) |
| Website | • Harris et al [76]<br>• Frisch et al [78]<br>• Nguyen et al [80]<br>• Perrault and Silk [81] | 5 (14) |
| Other social media (YouTube and Myspace) | • Potente et al [27] | 1 (3) |

### *Engagement*

In this group, 54% (19/35) of the studies analyzed the engagement of campaigns or interventions aimed at promoting disease prevention or health promotion measures.

One of the main takeaways from these studies is that the message format affects the level of users' web-based commitment to health organizations via social media [51]. Other studies (2/35, 6%) were tied to the assessment of the effectiveness of different types of posts on Facebook. A study analyzed the extent to which a post can resonate with an indecisive parent when it comes to the human papillomavirus vaccine [52], concluding that designing factual posts so that they include an emotional dimension increased the engagement with these posts, sponsored content can generate more negative comments than organic content, and all people should be addressed in an accommodating manner regardless of their tone. Although Syred et al [53] stated that moderation can help maintain the discussion quality and generate new interest and discussion on a certain topic, Loft et al [52], by contrast, focused more on the technical aspects of Facebook posts by stating that photos with short comments were the most effective in engaging information consumers and greater use of this post type could encourage greater audience engagement. At the same time, professional videos may not be as effective as a mechanism for active audience engagement on social media platforms. Zhang and Zhou [54] analyzed message efficacy, this time with particular attention to fear, and proposed a strategy where the inclusion of more emotional cues such as pictures is emphasized to arouse fear to motivate information dissemination on social media. In the social media context (in this case, Twitter and anonymous web-based discussion platforms), the relative importance of having clear informational content sent from organizations would be much greater for generating highly viewed and shared cancer prevention messages [55].

The studies by Kite et al [56], Lister et al [57], Paracka1 et al [58], Rabarison et al [59], and Reuter et al [60] measured engagement in and of itself. In particular, the study by Rabarison et al [59] focused on a specific aspect of the social media in question: Twitter chats. Specifically, chats of this kind should be used as an engagement tool with the audience by sharing messages and responding to questions from the public. Focusing on Twitter, Instagram, and Facebook, Reuter et al [60] concluded that engagement with a health message on social media does not indicate user engagement on a website and, therefore, it is recommended that both metrics be taken into account when designing health promotion strategies. It was also suggested to combine organic and advertising messages in health promotion campaigns. More specifically, with regard to Facebook, communication effectiveness could be enhanced using a two-way communication format, which enables the promoter to respond to negative comments [58]. Finally, according to the studies by Kite et al [56] and Lister et al [57], effective engagement through Facebook requires both maximizing the reach of posts through paid boosts and delivering content that users want to engage with and share to capitalize on word-of-mouth marketing.

In total, 6% (2/35) of the studies suggested social media strategies with the aim of improving engagement by comparing the work of institutional social media with that of lifestyle influencers [61] or by investigating the way in which network structures explain retweeting behaviors [62]. More specifically, Klassen et al [63] and Kim et al [62] stated, respectively, that health promotion organizations should try to build relationships with their users in a similar fashion to lifestyle brands and that influential people should be identified and targeted as their messages are more likely to be disseminated.

Regarding more technical aspects, the study by Guidry et al [64] states that, in the case of crisis communication, public health organizations should be present on all major social media platforms, but Instagram may yield the greatest return and user engagement. The study by Chung [65] was aimed at examining whether dialogic messages induced greater risk-preventive behavioral intentions than monologic messages, reaching the conclusion that frequent posting of tweets with images and graphs instead of videos and hyperlinks is beneficial. Similarly, according to Barklamb et al [61], strategies that were associated with higher engagement included the use of hashtags and announcements compared with not prompting engagement strategies. However, imagery should be carefully used as it appeared to be a powerful tool for attracting attention and briefly engaging users (ie, increasing likes) as well as increasing message transmission (ie, increasing shares). However, the use of images with information about illness consequences and control or with messages conveying negative affect could mute responses [66]. In particular, communication effectiveness could also be enhanced by designing factual posts so that they include an emotional dimension that could increase engagement [67]. Moreover, according to Strekalova and Krieger [67], sponsored content can generate more negative comments than organic content, and all people should be addressed in an accommodating manner regardless of their tone. Finally, users were more likely to click, share, comment, or like the content of posts that had photos. Branded, visual content was more effective in facilitating engagement [68].

Finally, 3% (1/35) of the studies focused on the effectiveness of communication theories, in particular that of the health belief model. Analyzing the Instagram accounts of the health departments of the Gulf Cooperation Council, it was found that the health belief model should be included more in internet-based communication [69].

### *Message Framing*

Considerably less studies (8/35, 23%) dealt with the topic of message framing in this category. The first study in this category was by Dockter et al [24], stating that content should be transmitted or retransmitted by well-known, credible sources. On a more specific note relating to content engagement, Yoo et al [70] recommended the use of content-oriented social media when trying to influence risk perception during campaigns, with particular attention to posts with photos as users were more likely to click, share, comment, or like this type of content. Borah and Xiao [28] and Cho et al [71] investigated the effect of health message framing and the moderating effects of social endorsement and source type on credibility perceptions of posts, resulting in a superiority of gain-framed messages to reach a positive campaign outcome. Other studies (4/35, 11%) were tied to the assessment of the effectiveness of different types of posts on Facebook when it comes to engagement [72]. In particular, communication effectiveness could be enhanced using a two-way communication format, which enables the promoter to respond to negative comments. Nobles et al [73] examined the demographic profile in photos concerning HIV prevention and diagnosis, underlining a disparity in the representation of minorities and marginalized communities. Another study by Chung and Lim [74] focused on a long-running campaign on National Breast Cancer Awareness Month and concluded with 2 observations regarding the efficacy of frequent posting and the positive impact of photos and images instead of videos and hyperlinks. Finally, Whitten et al [75] addressed the presence of information targeting low-literacy, racially diverse, non–English-speaking, and age-diverse audiences on breast cancer websites. The results were three-fold: if content were tagged according to ethnicity or language, then this would allow users to browse websites according to the information that is most personally relevant; it would be beneficial for websites containing lower-literacy material to avoid statistical data; and storytelling evidence has demonstrated the ability to serve as a greater motivator for healthy behaviors.

### *Campaign Effectiveness*

This last group included 23% (8/35) of the studies, which focused on the assessment of the success of a particular campaign. Starting again from a more general framework, the study by Harris et al [76] stated that there are 4 qualities that are key to influencing trust and the subsequent decision to act on the advice given. These are information quality, personalization, perceived impartiality, and design credibility. Delving deeply into the issue of trust, the studies described in this section proposed different strategies to maximize trust from the web users. Social media can also be used effectively in social marketing campaigns and is an essential tool in the promotional mix when targeting young people. According to Potente et al [27], entertaining peer-to-peer messages can be used to engage youths with an important health message for skin cancer prevention. By contrast, Allen et al [77], on the promotion of the human papillomavirus vaccine, recorded no statistically significant change in the intent to be vaccinated in the next 6 or 12 months after the campaign among those who had not yet started or completed vaccination. Focusing on more technical aspects, Frisch et al [78] stated that websites designed for health education should include visual presentations of information such as pictures, charts, or graphs. Similarly, Yoo et al [79] were concerned with more technical aspects of Twitter communication, especially when developing a Twitter campaign. The results included the need to consider incorporating features such as hyperlinks to related websites or live chats with health care providers as well as the creation of tailored messages or edutainment, which may also be considered to engage people in the process of information selection and transmission. Moving forward from just design structure, the way content is presented is also a great source of studies. In this case, Nguyen et al [80] offered useful insights, concluding that mode tailoring may be a tool to reduce or prevent the information overload that may occur when too much information is placed on a nontailored web page at one time. Other than content, the effectiveness of website-based communication is measured through ease of navigation and trust in the information provided by the website. Perrault and Silk [81] used social cognitive theory and media richness theory to prove that the exposure to videos was responsible for the increased engagement in risk-reduction behaviors. Another communication theory is the transtheoretical model. This was used in the study by Pirzadeh et al [82], who stated that the transtheoretical model was the most effective education strategy when it comes to prompting behavior change.

## Studies on General Health

### *Overview*

In the category of general health (Table 3), 17% (13/78) of the studies were included. Of these 13 studies, 9 (69%) were carried out in the United States, whereas 3 (23%) were carried out in the United Kingdom. Engagement was, once again, the most represented topic (9/13, 69%), followed by message framing (4/13, 31%). The same proportion holds true when it comes to studies on collections of posts (9/13, 69%) versus studies on human samples (4/13, 31%). The study designs were not overly diversified in this group as 46% (6/13) were observational studies and 38% (5/13) were content analyses. In this group, all studies (13/13, 100%) were conducted on a single communication medium, but only 31% (4/13) were connected with a specific communication theory (the remaining 9/13, 69% were not). The Kmet evaluation resulted in an average of 0.76 points, with a correlation coefficient of 0.82.

**Table 3.** List of studies on general health (see Multimedia Appendix 1 for more details; N=13).

| Primary evaluated aspect and communication medium | Reference | Studies, n (%) |
|---|---|---|
| **Engagement** | | |
| Facebook | • Bhattacharya et al [86]<br>• Kite et al [87]<br>• Campbell and Rudan [88] | 3 (23) |
| Twitter | • Bhattacharya et al [83]<br>• Fung et al [84]<br>• Meng et al [85]<br>• Park et al [90] | 4 (31) |
| Instagram | • Kim and Kim [89] | 1 (8) |
| Website | • Pang et al [91]<br>• Hung and Stones [92]<br>• Lazard and Mackert [93]<br>• Shim and Jo [94]<br>• Sillence et al [95] | 5 (38) |

### *Engagement*

In this group, 15% (2/13) of the studies [83,84] focused on the content of health organizations' Twitter profiles, concluding that the use of hashtags, URLs, visual cues, and user mentions was positively associated with retweets. Another study focusing on retweets and on the type of message brought on by health organizations is the one by Meng et al [85]. They defined a series of recommendations on the matter (ie, designing efficacious information is the key to increasing the aggregated number of retweets, crafting information that can raise risk perception is important to increase the diffusion chain through person-to-person transmission, and tweets that induce negative emotions could be more effective in catching users' attention and expanding sharing of the information). A similar study by Bhattacharya et al [86], this time on Facebook, also stated that posts containing media or links and expressing positive sentiments correlated with higher or longer engagement. Facebook was also the topic of research of Kite et al [87], who concluded that content providers must encourage engagement and adapt to the Facebook algorithm to maximize message exposure while also ensuring that the content is of high quality. Language also plays an important role in the effectiveness of a post, as explained by Campbell and Rudan [88], who claimed that adjusting the language and presentation can be of more appeal to popular culture. Kim and Kim [89], by contrast, centered their study on the Instagram presence of the Centers for Disease Control and Prevention (CDC), stating that the message design should be different depending on whether the aim is to increase the number of likes and comments or induce a more positive response from the public. The dissemination of health information is also a topical area of research, in particular by Park et al [90], who provided guidelines such as retweeting content from health information sources with a high number of Twitter followers to build up an organization's follower base. The study by Pang et al [91] was aimed at creating a design for a consumer health website by supporting different health-seeking behaviors. In particular, different types of information-seeking behavior should be supported as access to a dynamic information scope is critical for health information seeking.

### *Message Framing*

This group comprised 4 studies: the studies by Hung and Stones [92], Lazard and Mackert [93], Shim and Jo [94], and Sillence et al [95].

Concerning design and website presentation, Lazard and Mackert [93] stated that high design complexity is often associated with a better perceived comprehensibility, a greater perceived usefulness, a greater message quality, and higher perceptions of visual informativeness. Other than content, the effectiveness of website-based communication is measured through ease of navigation and trust in the information provided by the website. Sillence et al [95] identified the key factors influencing UK and US citizens' trust and intention to act on advice found on health websites (ie, credibility and impartiality). Moving forward from just design structure, the way content is presented is also a great source of studies. In this case, Hung and Stones [92] offered useful insights, stating that, among other guidelines, to appeal to the public, general terms should substitute professional terms and simplified text-based content should be used. Delving deeply into the issue of trust, Shim and Jo [94] applied the information systems success model, assessing that service quality had a significant association with user satisfaction and that its impact on perceived benefits occurred indirectly to user satisfaction and intention, thus maximizing trust from the web users.

## Health Literacy and Misinformation Correction

This is the smallest group in this corpus (Table 4), with only 8% (6/78) of the studies. They are divided into 2 categories: misinformation correction (4/6, 67%) and health literacy (2/6, 33%). Given the scarcity of examples of these types of interventions, they were grouped together. Of these 6 studies, 4 (67%) were carried out in the United States. Contrary to the other groups, studies focused on the analysis of engagement were not present in this group, and only message framing and campaign effectiveness were represented. All the studies (6/6, 100%) had groups of participants as their main sample. All the studies in this group (6/6, 100%) referred to a single communication medium. Of the 6 studies, 1 (17%) did not present a reference to a specific communication theory. The Kmet evaluation resulted in an average of 0.80 points, with a correlation coefficient of 0.59.

The misinformation correction studies were those by Bode and Vraga [96], Gesser-Edelsburg et al [97], and Vraga and Bode [98,99]. First, Bode and Vraga [96] stated that corrective information provided via an algorithm or social connections reduces misinformation and is effective as they are considered credible, whereas Gesser-Edelsburg et al [97] stated that it is important for organizations to correct misinformation transparently while at the same time addressing the emotional aspects that may come into play in case of conflicts of opinion. The study by Vraga and Bode [99] was carried out to test the efficacy of shareable infographics to debunk COVID-19 myths. In particular, one of the solutions found by the authors was that organizations can debunk misinformation circulating in society by sharing high-quality information on social media, emphasizing the facts without waiting to see them shared directly in their feeds, which expands the opportunities for observational correction to occur. Information correction is an area of interest that is also reported in this group, with the study by Vraga and Bode [98] testing whether the number and source (user vs the CDC) of corrective responses affect the successful reduction of misperceptions. Thus, this study suggests that organizations should speak up when they see misinformation on social media and reputable organizations such as the CDC should monitor social media feeds and immediately attempt to rebut misinformation when it arises.

The last 33% (2/6) of the studies dealt with health literacy and were all designed to develop different approaches aimed at different categories. The first of these studies is the one by Chin et al [100], which focused on older adults and proposed a multifaceted approach guided by theories of text comprehension and document design to improve readability for websites, in particular claiming that increasing document readability alone is insufficient for improving text comprehension in older adults. Meppelink et al [101] added to this statement by suggesting the use of a visual representation of information to improve the retention of information.

**Table 4.** List of studies on misinformation correction and health literacy (see Multimedia Appendix 1 for more details; N=6).

| Primary evaluated aspect and communication medium | Reference | Studies, n (%) |
|---|---|---|
| **Message framing** | | |
| Facebook | - Gesser-Edelsburg et al [97]<br>- Vraga and Bode [99] | 2 (33) |
| Website | - Chin et al [100] | 1 (17) |
| **Campaign effectiveness** | | |
| Facebook | - Bode and Vraga [96] | 1 (17) |
| Twitter | - Vraga and Bode [98] | 1 (17) |
| Website | - Meppelink et al [101] | 1 (17) |

## Quality Assessment

The Kmet evaluation was used to distinguish between studies presenting a solid structure and studies lacking those factors, as made explicit by the low score obtained for the Kmet evaluation. In particular, considering the 0.75-point threshold, 36% (28/78) of the studies were excluded. To be more specific, of the 78 studies, 3 (4%) had <0.55 points, 25 (32%) were between the 0.55- and the 0.75-point mark, and 50 (64%) were above the 0.75-point threshold.

Table 5 shows the distribution of the studies' quality for each research category using the 0.75-point mark as a threshold. The health promotion and disease prevention category had a higher percentage of good-quality studies than the other categories (*P*=.02).

**Table 5.** Research categories and distribution of quality studies according to the 0.75-point Kmet score threshold (N=78).

| | Studies below cutoff, n (%) | Studies above cutoff, n (%) |
|---|---|---|
| Crisis communication (n=24) | 13 (54) | 11 (46) |
| Health promotion and disease prevention (n=35) | 10 (29) | 25 (71) |
| General health (n=13) | 9 (69) | 4 (31) |
| Misinformation correction and health literacy (n=6) | 1 (17) | 5 (83) |

# Discussion

## Principal Findings

Our review provides insights into topics regarding the different modes of communication used by health care authorities to engage with the public in different situations, namely, crisis communication and health promotion and disease prevention. Previous systematic reviews have dealt with this topic by focusing on certain specific aspects such as communication effectiveness for specific channels and situations. This systematic review aimed to provide a more comprehensive view of internet-based health communication. The amount of works included in this study also suggests a high interest in this particular topic. It is also worth mentioning that health communication represents a vital point for public health as the rapid diffusion of information to the largest possible number of users is key when trying to effectively communicate important information, as also recently seen during the COVID-19 pandemic.

The results of this systematic review raise an important question: is it possible to define a series of key points to address the basics of internet-based communication for public health?

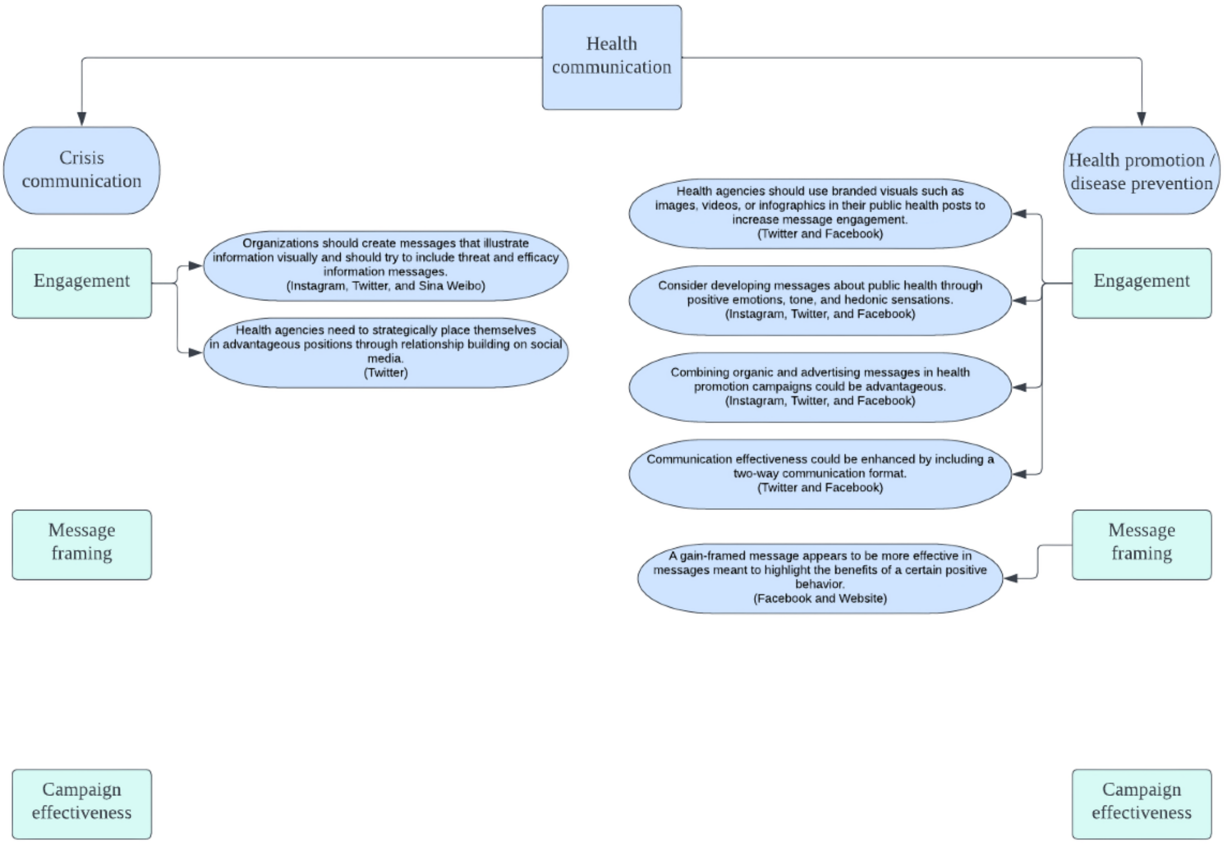
To do so, a flowchart (Figure 2) was created, starting from the basic distinction between the 2 main themes that are addressed by the selected studies: crisis communication and health promotion and disease prevention. The other 2 categories identified in this review (general health and health literacy and misinformation correction) were not considered as they were of too general or too narrow scope to constitute a relevant sample. This distinction was made necessary as these 2 types of communication account for 2 almost opposite situations and purposes.

Going further into this analysis, it was vital to make a distinction between the different primary evaluated aspects (ie, engagement, message framing, and campaign effectiveness). This was done because the conventions and communication strategies used differed greatly, especially given the different nature of communication for those purposes. An important point to be made regards the criteria according to which the key points were chosen to be included in this flowchart. A first measure is represented by the Kmet score of the study. As this score is used to address the overall soundness of the research process, only the studies that recorded a score ≥0.75 points were taken into account. Another measure included was the repetition of certain suggested recommendations or conclusions throughout the group in question. What this entails is that a certain specific proposed strategy that was repeated at least two times was included in the final flowchart as it can be assumed that it was more easily applicable in a more general context. Figure 2 offers a more thorough rundown of the included key points identified in this systematic review. As we can see, not all the primary evaluated aspects are paired with one or more key points as, in some cases, the studies in question did not meet the selected criteria. As for crisis communication, only the engagement category presented 2 key points: one regarding the need to create messages that visually convey information and the other addressing the need for health agencies to place themselves in advantageous positions when it comes to relationship building on social media. As for health promotion and disease prevention, engagement recorded 4 key points having to do with creating effective visual information, promoting the use of a positive tone in messages, combining organic and advertising messages, and implementing a two-way communication. One last indication was made for message framing, underlying the greater effectiveness of gain-framed communication.

To be noted is also the fact that the key points proposed are not universally applicable to all communication channels but, rather, to specific ones. The proposed flowchart includes indications as to which communication channel the proposed key points are applicable to (ie, websites, Facebook, Twitter, Instagram, and Sina Weibo).

This systematic review met the criteria of Assessing the Methodological Quality of Systematic Reviews [102], a measurement tool to assess the methodological quality of systematic reviews, thus ensuring the accuracy of the reviewed data. However, some limitations should be addressed.

**Figure 2.** Flowchart of key points.



## Limitations

Through a systematic approach, we aimed to provide a comprehensive look at health care communication via different media and with different purposes. Although the number of articles retrieved was relatively large, some limitations related to the study design must be underlined. As already mentioned, the large number of studies not meeting the threshold (28/78, 36%) suggests a lack of soundness of the selected studies, thus calling attention to a need for more in-depth research on the topic of internet-based communication as well as measures for campaign and intervention effectiveness.

This need is also tied to another limitation of this study, posed this time by the lack of appropriate measures to evaluate the quality of the studies of this sort. In this case, the Kmet evaluation tool, albeit flawed, represented the best possible measure to evaluate studies on health communication. However, this tool is directed at the assessment of primary research reports in the field of medical experimental research, and a number of the areas of evaluation were impossible to relate to these studies (ie, randomization, double-blind, sample size appropriateness, and control for confounding). As the scale for evaluation appears to be rather limited, the scores attributed to the studies in this systematic review were based on an average of circa 18 maximum points against an actual maximum of 24 points. Therefore, this raises the need for a more precise tool to evaluate this type of studies.

Another issue lies in the fact that the studies taken into account for this systematic review do not have a uniform end point as, for example, some focus on the reach of a campaign or on user behaviors on social media. This results in a multitude of different measures of success that make it difficult to properly understand and evaluate the reach and success of a campaign or of certain web-based behaviors. This variety in end points is also reflected in the variety—or lack thereof—of measures of success of a certain campaign or policy. Thus, it is difficult to assess a baseline measure of effectiveness for each of the communication channels described in this review, which points to the necessity of forming a medium-specific criterion for this kind of evaluation.

A final limitation is related to the extreme specificity of web-based communication based on geographical as well as cultural differences that make it difficult to form a comprehensive list of guidelines for this type of discourse. The use of the internet and social media and the strategies and practices adopted by single countries or even smaller cultural groups is an issue so big that it is impossible to look away from it when considering this type of studies, which makes it difficult to redact a list of guidelines to adopt when trying to manage internet-based health communication. Thus, this creates the need to always address the country of origin of a study as well as its specific target audience.

## Conclusions

The evidence gathered in this study suggests that no single strategy works best in the case of health care communication. Although there is evidence supporting multiple communication approaches across different media, how the interaction unfolds must be resolved according to a number of variables: communication medium (website or social media), country of intervention, organization type (health organization or health ministry), and aim of the intervention.

This extreme variability of outcomes and the lack of a unitary measure for assessing the end points of a specific campaign or study on individuals lies in the inherently fluid and ever-changing essence of communication practices, which makes it difficult to define this concept altogether as well as grasp a precise definition of what *evidence* entails in this field compared, for example, with the evidence gathered through randomized controlled trials and cohort studies in the medical field.

In practice, this review tried to provide a baseline for practitioners and researchers as to how to conduct a campaign on the web on different web-based communication channels. However, as stated before, this is not enough to provide a comprehensive set of guidelines on the matter; rather, as a matter of fact, it raises more questions that need to be addressed in future research, in particular on the matter of forming a unified measure of effectiveness for campaigns and policies and on the scales used to evaluate the soundness of a certain study.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Extended tables describing the studies collected for this research.
[PDF File (Adobe PDF File), 1047 KB - jmir_v24i9e38541_app1.pdf ]

## References

1. Sandhu S. Strategic communication: an institutional perspective. Int J Strategic Commun 2009 Apr 10;3(2):72-92. [doi: 10.1080/15531180902805429]
2. Hallahan K, Holtzhausen D, van Ruler B, Verčič D, Sriramesh K. Defining strategic communication. Int J Strategic Commun 2007 Mar 22;1(1):3-35. [doi: 10.1080/15531180701285244]
3. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: an empirical analysis of a national cancer-related survey. Decision Support Syst 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
4. Global Digital population as of January 2021. Statista. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2021-11-17]
5. Bennett GG, Glasgow R. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
6. Gesser-Edelsburg A. Using narrative evidence to convey health information on social media: the case of COVID-19. J Med Internet Res 2021 Mar 15;23(3):e24948 [FREE Full text] [doi: 10.2196/24948] [Medline: 33674257]
7. van Velsen L, van Gemert-Pijnen JE, Beaujean DJ, Wentzel J, van Steenbergen JE. Should health organizations use web 2.0 media in times of an infectious disease crisis? An in-depth qualitative study of citizens' information behavior during an EHEC outbreak. J Med Internet Res 2012 Dec 20;14(6):e181 [FREE Full text] [doi: 10.2196/jmir.2123] [Medline: 23257066]
8. WHO Strategic Communications Framework for effective communication. World Health Organization. 2017. URL: https://tinyurl.com/383phdua [accessed 2021-12-02]
9. Chong YY, Cheng HY, Chan HY, Chien WT, Wong SY. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
10. Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. J Public Health Policy 2020 Dec 21;41(4):410-420 [FREE Full text] [doi: 10.1057/s41271-020-00247-w] [Medline: 32826935]
11. Berg S, O'Hara JK, Shortt M, Thune H, Brønnick KK, Lungu DA, et al. Health authorities' health risk communication with the public during pandemics: a rapid scoping review. BMC Public Health 2021 Jul 15;21(1):1401 [FREE Full text] [doi: 10.1186/s12889-021-11468-3] [Medline: 34266403]
12. Alamoodi A, Zaidan B, Zaidan A, Albahri O, Mohammed K, Malik RQ, et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: a systematic review. Expert Syst Appl 2021 Apr 01;167:114155 [FREE Full text] [doi: 10.1016/j.eswa.2020.114155] [Medline: 33139966]
13. Fung I, Blankenship E, Ahweyevu J, Cooper LK, Duke CH, Carswell S, et al. Public health implications of image-based social media: a systematic review of Instagram, Pinterest, Tumblr, and Flickr. Permanente J 2020 Mar;24(1). [doi: 10.7812/tpp/18.307]
14. Giustini D, Ali S, Fraser M, Kamel Boulos MN. Effective uses of social media in public health and medicine: a systematic review of systematic reviews. Online J Public Health Inform 2018;10(2):e215 [FREE Full text] [doi: 10.5210/ojphi.v10i2.8270] [Medline: 30349633]
15. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016 Jun;22(2):355-369 [FREE Full text] [doi: 10.1177/1460458214559432] [Medline: 25518944]
16. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011 Jul 22;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]

17. Lim M, Wright C, Carrotte E, Pedrana A. Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites. Health Promot J Austr 2016 Oct 06;27(3):187-197. [doi: 10.1071/he16057]
18. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
19. 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]
20. Fernandez P. Zotero: information management software 2.0. Library Hi Tech News 2011 Jun 07;28(4):5-7. [doi: 10.1108/07419051111154758]
21. Puckett J. Zotero A Guide for Librarians, Researchers, and Educators. Chicago, Illinois, United States: Association of College and Research Libraries; 2011.
22. Kmet L, Cook L, Lee R, Alberta Heritage Foundation for Medical Research. Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields. Alberta, Canada: Alberta Heritage Foundation for Medical Research; 2004.
23. An introduction to risk communication. World Health Organization. 2014 Aug 11. URL: https://www.who.int/publications/i/item/an-introduction-to-risk-communication [accessed 2022-01-20]
24. Dockter CE, Lee S, Boman CD, Hinnant A, Cameron GT. The impact of retransmission and modality on communicating health research findings via social media. Health Commun 2021 Sep 09;36(10):1231-1241. [doi: 10.1080/10410236.2020.1749354] [Medline: 32268798]
25. Kim H, Saffer A, Liu W, Sun J, Li Y, Zhen L, et al. How public health agencies break through COVID-19 conversations: a strategic network approach to public engagement. Health Commun 2022 Sep;37(10):1276-1284. [doi: 10.1080/10410236.2021.1886393] [Medline: 33591839]
26. Chen Q, Min C, Zhang W, Ma X, Evans R. Factors driving citizen engagement with government TikTok accounts during the COVID-19 pandemic: model development and analysis. J Med Internet Res 2021 Feb 04;23(2):e21463 [FREE Full text] [doi: 10.2196/21463] [Medline: 33481756]
27. Potente S, McIver J, Anderson C, Coppa K. “it's a beautiful day for cancer”: an innovative communication strategy to engage youth in skin cancer prevention. Social Marketing Q 2011 Sep 01;17(3):86-105. [doi: 10.1080/15245004.2011.595604]
28. Borah P, Xiao X. The importance of 'likes': the interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. J Health Commun 2018;23(4):399-411. [doi: 10.1080/10810730.2018.1455770] [Medline: 29601271]
29. MacKay M, Colangeli T, Gillis D, McWhirter J, Papadopoulos A. Examining social media crisis communication during early COVID-19 from public health and news media for quality, content, and corresponding public sentiment. Int J Environ Res Public Health 2021 Jul 28;18(15):7986 [FREE Full text] [doi: 10.3390/ijerph18157986] [Medline: 34360278]
30. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: how to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]
31. Dimanlig-Cruz S, Han A, Lancione S, Dewidar O, Podinic I, OPH-SEPH Collaborative (OSC) Working Group, et al. Physical distancing messages targeting youth on the social media accounts of Canadian public health entities and the use of behavioral change techniques. BMC Public Health 2021 Sep 07;21(1):1634 [FREE Full text] [doi: 10.1186/s12889-021-11659-y] [Medline: 34493258]
32. Hagen L, Keller T, Neely S, DePaula N, Robert-Cooperman C. Crisis communications in the age of social media. Social Sci Comput Rev 2017 Aug 21;36(5):523-541. [doi: 10.1177/0894439317721985]
33. Lauran N, Kunneman F, Van de Wijngaert L. Connecting social media data and crisis communication theory: a case study on the chicken and the egg. J Risk Res 2019 Jul 03;23(10):1259-1277. [doi: 10.1080/13669877.2019.1628097]
34. Slavik C, Buttle C, Sturrock S, Darlington J, Yiannakoulias N. Examining tweet content and engagement of Canadian public health agencies and decision makers during COVID-19: mixed methods analysis. J Med Internet Res 2021 Mar 11;23(3):e24883 [FREE Full text] [doi: 10.2196/24883] [Medline: 33651705]
35. McInnes C. Add twitter and stir’: the use of twitter by public authorities in Norway and UK during the 2014-15 Ebola outbreak. Observatorio 2018;12(2). [doi: 10.15847/obsobs12220181173]
36. Vos S, Sutton J, Yu Y, Renshaw SL, Olson MK, Gibson CB, et al. Retweeting risk communication: the role of threat and efficacy. Risk Anal 2018 Dec;38(12):2580-2598. [doi: 10.1111/risa.13140] [Medline: 30080933]
37. Lwin M, Lu J, Sheldenkar A, Schulz P. Strategic uses of Facebook in Zika outbreak communication: implications for the crisis and emergency risk communication model. Int J Environ Res Public Health 2018 Sep 10;15(9):1974 [FREE Full text] [doi: 10.3390/ijerph15091974] [Medline: 30201929]
38. Sutton J, Renshaw S, Butts C. COVID-19: retransmission of official communications in an emerging pandemic. PLoS One 2020;15(9):e0238491 [FREE Full text] [doi: 10.1371/journal.pone.0238491] [Medline: 32936804]

39. Tang L, Liu W, Thomas B, Tran H, Zou W, Zhang X, et al. Texas public agencies' tweets and public engagement during the COVID-19 pandemic: natural language processing approach. JMIR Public Health Surveill 2021 Apr 26;7(4):e26720 [FREE Full text] [doi: 10.2196/26720] [Medline: 33847587]
40. Young R, Tully M, Dalrymple K. #Engagement: use of Twitter chats to construct nominal participatory spaces during health crises. Inform Commun Soc 2017 Mar 27;21(4):499-515. [doi: 10.1080/1369118x.2017.1301518]
41. Yang Y, Deng W, Zhang Y, Mao Z. Promoting public engagement during the COVID-19 crisis: how effective is the Wuhan local government's information release? Int J Environ Res Public Health 2020 Dec 26;18(1):118 [FREE Full text] [doi: 10.3390/ijerph18010118] [Medline: 33375307]
42. Guidry J, Jin Y, Orr C, Messner M, Meganck S. Ebola on Instagram and Twitter: how health organizations address the health crisis in their social media engagement. Public Relation Rev 2017 Sep;43(3):477-486. [doi: 10.1016/j.pubrev.2017.04.009]
43. Renshaw S, Mai S, Dubois E, Sutton J, Butts C. Cutting through the noise: predictors of successful online message retransmission in the first 8 months of the COVID-19 pandemic. Health Secur 2021;19(1):31-43 [FREE Full text] [doi: 10.1089/hs.2020.0200] [Medline: 33606574]
44. Vos SC, Sutton J, Gibson CB, Butts CT. #Ebola: emergency risk messages on social media. Health Secur 2020 Dec;18(6):461-472. [doi: 10.1089/hs.2019.0158] [Medline: 33326333]
45. Pascual-Ferrá P, Alperstein N, Barnett D. Social network analysis of COVID-19 public discourse on twitter: implications for risk communication. Disaster Med Public Health Prep 2022 Apr;16(2):561-569 [FREE Full text] [doi: 10.1017/dmp.2020.347] [Medline: 32907685]
46. Ort A, Fahr A. Using efficacy cues in persuasive health communication is more effective than employing threats - An experimental study of a vaccination intervention against Ebola. Br J Health Psychol 2018 Sep 10;23(3):665-684. [doi: 10.1111/bjhp.12310] [Medline: 29635864]
47. Jang K, Baek Y. When information from public health officials is untrustworthy: the use of online news, interpersonal networks, and social media during the MERS outbreak in South Korea. Health Commun 2019 Aug;34(9):991-998. [doi: 10.1080/10410236.2018.1449552] [Medline: 29558170]
48. Sutton J, Renshaw S, Butts C. The first 60 days: American public health agencies' social media strategies in the emerging COVID-19 pandemic. Health Secur 2020 Dec;18(6):454-460. [doi: 10.1089/hs.2020.0105] [Medline: 33047982]
49. Duong H, Nguyen L, Julian McFarlane S, Nguyen H, Nguyen K. Preventing the COVID-19 outbreak in Vietnam: social media campaign exposure and the role of interpersonal communication. Health Commun 2021 Jul 19:1-8. [doi: 10.1080/10410236.2021.1953729] [Medline: 34278892]
50. Harris-Sagaribay L, Chambers C, Perrotta K, Polen K, Honein M, Wasternack E. A model partnership for communication and dissemination of scientific recommendations for pregnant women during the emergency response to the Zika virus outbreak: MotherToBaby and the Centers for Disease Control and Prevention. Birth Defects Res 2020 Nov;112(18):1545-1550 [FREE Full text] [doi: 10.1002/bdr2.1787] [Medline: 32840950]
51. Alonso-Cañadas J, Galán-Valdivieso F, Saraite-Sariene L, Caba-Pérez C. Committed to health: key factors to improve users' online engagement through Facebook. Int J Environ Res Public Health 2020 Mar 11;17(6):1814 [FREE Full text] [doi: 10.3390/ijerph17061814] [Medline: 32168792]
52. Loft L, Pedersen E, Jacobsen S, Søborg B, Bigaard J. Using Facebook to increase coverage of HPV vaccination among Danish girls: an assessment of a Danish social media campaign. Vaccine 2020 Jun 26;38(31):4901-4908. [doi: 10.1016/j.vaccine.2020.04.032] [Medline: 32362529]
53. Syred J, Naidoo C, Woodhall SC, Baraitser P. Would you tell everyone this? Facebook conversations as health promotion interventions. J Med Internet Res 2014 Apr 11;16(4):e108 [FREE Full text] [doi: 10.2196/jmir.3231] [Medline: 24727742]
54. Zhang X, Zhou S. Sharing health risk messages on social media: effects of fear appeal message and image promotion. Cyberpsychology 2020 May 13;14(2). [doi: 10.5817/cp2020-2-4]
55. Zhang J, Le G, Larochelle D, Pasick R, Sawaya G, Sarkar U, et al. Facts or stories? How to use social media for cervical cancer prevention: a multi-method study of the effects of sender type and content type on increased message sharing. Prev Med 2019 Sep;126:105751 [FREE Full text] [doi: 10.1016/j.ypmed.2019.105751] [Medline: 31226342]
56. Kite J, Grunseit A, Li V, Vineburg J, Berton N, Bauman A, et al. Generating engagement on the make healthy normal campaign Facebook page: analysis of Facebook analytics. JMIR Public Health Surveill 2019 Jan 14;5(1):e11132 [FREE Full text] [doi: 10.2196/11132] [Medline: 31344679]
57. Lister C, Royne M, Payne H, Cannon B, Hanson C, Barnes M. The laugh model: reframing and rebranding public health through social media. Am J Public Health 2015 Nov;105(11):2245-2251. [doi: 10.2105/ajph.2015.302669]
58. Parackal M, Parackal S, Eusebius S, Mather D. The use of Facebook advertising for communicating public health messages: a campaign against drinking during pregnancy in New Zealand. JMIR Public Health Surveill 2017 Aug 10;3(3):e49 [FREE Full text] [doi: 10.2196/publichealth.7032] [Medline: 28798011]
59. Rabarison K, Croston M, Englar N, Bish C, Flynn S, Johnson C. Measuring audience engagement for public health Twitter chats: insights from #LiveFitNOLA. JMIR Public Health Surveill 2017 Jun 08;3(2):e34 [FREE Full text] [doi: 10.2196/publichealth.7181] [Medline: 28596149]

60. Reuter K, Wilson M, Moran M, Le N, Angyan P, Majmundar A, et al. General audience engagement with antismoking public health messages across multiple social media sites: comparative analysis. JMIR Public Health Surveill 2021 Feb 19;7(2):e24429. [doi: 10.2196/24429]
61. Barklamb A, Molenaar A, Brennan L, Evans S, Choong J, Herron E, et al. Learning the language of social media: a comparison of engagement metrics and social media strategies used by food and nutrition-related social media accounts. Nutrients 2020 Sep 16;12(9):2839 [FREE Full text] [doi: 10.3390/nu12092839] [Medline: 32948033]
62. Kim E, Hou J, Han J, Himelboim I. Predicting retweeting behavior on breast cancer social networks: network and content characteristics. J Health Commun 2016;21(4):479-486. [doi: 10.1080/10810730.2015.1103326] [Medline: 27007166]
63. Klassen K, Borleis E, Brennan L, Reid M, McCaffrey T, Lim M. What people "Like": analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. J Med Internet Res 2018 Jun 14;20(6):e10227 [FREE Full text] [doi: 10.2196/10227] [Medline: 29903694]
64. Guidry J, Meganck S, Lovari A, Messner M, Medina-Messner V, Sherman S, et al. Tweeting about #diseases and #publichealth: communicating global health issues across nations. Health Commun 2020 Aug;35(9):1137-1145. [doi: 10.1080/10410236.2019.1620089] [Medline: 31131628]
65. Chung J. Retweeting in health promotion: analysis of tweets about Breast Cancer Awareness Month. Comput Human Behav 2017 Sep;74:112-119. [doi: 10.1016/j.chb.2017.04.025]
66. Rus H, Cameron L. Health communication in social media: message features predicting user engagement on diabetes-related facebook pages. Ann Behav Med 2016 Oct;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
67. Strekalova Y, Krieger J. A picture really is worth a thousand words: public engagement with the national cancer institute on social media. J Cancer Educ 2017 Mar;32(1):155-157 [FREE Full text] [doi: 10.1007/s13187-015-0901-5] [Medline: 26351003]
68. Theiss S, Burke R, Cory J, Fairley T. Getting beyond impressions: an evaluation of engagement with breast cancer-related Facebook content. Mhealth 2016 Nov;2:41 [FREE Full text] [doi: 10.21037/mhealth.2016.10.02] [Medline: 27840816]
69. Alkazemi MF, Guidry JP, Almutairi E, Messner M. #Arabhealth on Instagram: examining public health messages to Arabian Gulf State audiences. Health Commun 2022 Jan 02;37(1):39-47. [doi: 10.1080/10410236.2020.1816283] [Medline: 32873096]
70. Yoo W, Paek H, Hove T. Differential effects of content-oriented versus user-oriented social media on risk perceptions and behavioral intentions. Health Commun 2020 Jan 14;35(1):99-109. [doi: 10.1080/10410236.2018.1545169] [Medline: 30427203]
71. Cho J, Chun JW, Lee MJ. Impacts of message framing and social distance in health campaign for promoting regular physical exercise. J Health Commun 2018 Oct 16;23(9):824-835. [doi: 10.1080/10810730.2018.1527872] [Medline: 30325699]
72. Parackal M, Parackal S, Mather D, Eusebius S. Dynamic transactional model: a framework for communicating public health messages via social media. Perspect Public Health 2021 Sep;141(5):279-286. [doi: 10.1177/1757913920935910] [Medline: 32660353]
73. Nobles A, Leas E, Noar S, Dredze M, Latkin CA, Strathdee SA, et al. Automated image analysis of Instagram posts: implications for risk perception and communication in public health using a case study of #HIV. PLoS One 2020;15(5):e0231155 [FREE Full text] [doi: 10.1371/journal.pone.0231155] [Medline: 32365124]
74. Chung M, Lim Y. When health organization answers the question: differential effects of dialogic messages in website and Twitter through social presence and psychological distance. Health Commun 2022 May;37(6):685-695. [doi: 10.1080/10410236.2020.1864098] [Medline: 33356904]
75. Whitten P, Nazione S, Smith S, LaPlante C. An examination into audience targeting and the use of storytelling or statistical evidence on breast cancer websites. Patient Educ Couns 2011 Nov;85(2):e59-e64. [doi: 10.1016/j.pec.2011.01.008] [Medline: 21306857]
76. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]
77. Allen JD, Hollander J, Gualtieri L, Alarcon Falconi TM, Savir S, Agénor M. Feasibility of a twitter campaign to promote HPV vaccine uptake among racially/ethnically diverse young adult women living in public housing. BMC Public Health 2020 Jun 01;20(1):830 [FREE Full text] [doi: 10.1186/s12889-020-08824-0] [Medline: 32487045]
78. Frisch A, Camerini L, Schulz P. The impact of presentation style on the retention of online health information: a randomized-controlled experiment. Health Commun 2013;28(3):286-293. [doi: 10.1080/10410236.2012.683387] [Medline: 22716268]
79. Yoo S, Kim J, Lee Y. The effect of health beliefs, media perceptions, and communicative behaviors on health behavioral intention: an integrated health campaign model on social media. Health Commun 2018 Jan;33(1):32-40. [doi: 10.1080/10410236.2016.1242033] [Medline: 27858470]
80. Nguyen MH, van Weert JC, Bol N, Loos EF, Tytgat KM, van de Ven AW, et al. Tailoring the mode of information presentation: effects on younger and older adults' attention and recall of online information. Hum Commun Res 2016 Oct 21;43(1):102-126. [doi: 10.1111/hcre.12097]

81. Perrault E, Silk K. Testing the effects of the addition of videos to a website promoting environmental breast cancer risk reduction practices: are videos worth it? J Appl Commun Res 2014 Jan 01;42(1):20-40 [FREE Full text] [doi: 10.1080/00909882.2013.854400] [Medline: 25143661]
82. Pirzadeh A, Zamani F, Khoshali M, Kelishadi R. Web-based intervention on the promotion of physical activity among Iranian youth using the transtheoretical model. J Educ Health Promot 2020;9(1):118 [FREE Full text] [doi: 10.4103/jehp.jehp_36_20] [Medline: 32642474]
83. Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on twitter. PLoS One 2014 Nov 7;9(11):e112235 [FREE Full text] [doi: 10.1371/journal.pone.0112235] [Medline: 25379727]
84. Fung I, Jackson A, Mullican L, Blankenship EB, Goff ME, Guinn AJ, et al. Contents, Followers, and Retweets of the Centers for Disease Control and Prevention's Office of Advanced Molecular Detection (@CDC_AMD) Twitter Profile: cross-sectional study. JMIR Public Health Surveill 2018 Apr 02;4(2):e33 [FREE Full text] [doi: 10.2196/publichealth.8737] [Medline: 29610112]
85. Meng J, Peng W, Tan P, Liu W, Cheng Y, Bae A. Diffusion size and structural virality: the effects of message and network features on spreading health information on twitter. Comput Human Behav 2018 Dec;89:111-120 [FREE Full text] [doi: 10.1016/j.chb.2018.07.039] [Medline: 32288177]
86. Bhattacharya S, Srinivasan P, Polgreen P. Social media engagement analysis of U.S. Federal health agencies on Facebook. BMC Med Inform Decis Mak 2017 Apr 21;17(1):49 [FREE Full text] [doi: 10.1186/s12911-017-0447-z] [Medline: 28431582]
87. Kite J, Foley B, Grunseit A, Freeman B. Please Like Me: Facebook and public health communication. PLoS One 2016;11(9):e0162765 [FREE Full text] [doi: 10.1371/journal.pone.0162765] [Medline: 27632172]
88. Campbell I, Rudan I. Analysis of public engagement with ten major global health topics on a social network profile and a newspaper website. J Glob Health 2020 Jun;10(1):010902 [FREE Full text] [doi: 10.7189/jogh.10.010902] [Medline: 32257171]
89. Kim Y, Kim J. Using photos for public health communication: a computational analysis of the Centers for Disease Control and Prevention Instagram photos and public responses. Health Informatics J 2020 Sep;26(3):2159-2180 [FREE Full text] [doi: 10.1177/1460458219896673] [Medline: 31969051]
90. Park H, Reber B, Chon M. Tweeting as health communication: health organizations' use of Twitter for health promotion and public engagement. J Health Commun 2016;21(2):188-198. [doi: 10.1080/10810730.2015.1058435] [Medline: 26716546]
91. Pang P, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
92. Hung Y, Stones C. A comparative study of children’s ehealth design between east and west: a case study of children’s health websites in China, Taiwan, the UK, and the US. In: Ergonomics and Health Aspects of Work with Computers. Berlin, Heidelberg: Springer; 2011.
93. Lazard A, Mackert M. User evaluations of design complexity: the impact of visual perceptions for effective online health communication. Int J Med Inform 2014 Oct;83(10):726-735. [doi: 10.1016/j.ijmedinf.2014.06.010] [Medline: 25034279]
94. Shim M, Jo HS. What quality factors matter in enhancing the perceived benefits of online health information sites? Application of the updated DeLone and McLean Information Systems Success Model. Int J Med Inform 2020 May;137:104093. [doi: 10.1016/j.ijmedinf.2020.104093] [Medline: 32078918]
95. Sillence E, Blythe J, Briggs P, Moss M. A revised model of trust in internet-based health information and advice: cross-sectional questionnaire study. J Med Internet Res 2019 Nov 11;21(11):e11125 [FREE Full text] [doi: 10.2196/11125] [Medline: 31710297]
96. Bode L, Vraga E. See something, say something: correction of global health misinformation on social media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
97. Gesser-Edelsburg A, Diamant A, Hijazi R, Mesch G. Correcting misinformation by health organizations during measles outbreaks: a controlled experiment. PLoS One 2018;13(12):e0209505 [FREE Full text] [doi: 10.1371/journal.pone.0209505] [Medline: 30566485]
98. Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. Sci Commun 2017 Sep 14;39(5):621-645. [doi: 10.1177/1075547017731776]
99. Vraga EK, Bode L. Addressing COVID-19 misinformation on social media preemptively and responsively. Emerg Infect Dis 2021 Feb;27(2):396-403 [FREE Full text] [doi: 10.3201/eid2702.203139] [Medline: 33395379]
100. Chin J, Moeller D, Johnson J, Duwe EA, Graumlich JF, Murray MD, et al. A multi-faceted approach to promote comprehension of online health information among older adults. Gerontologist 2018 Jul 13;58(4):686-695 [FREE Full text] [doi: 10.1093/geront/gnw254] [Medline: 28329840]
101. Meppelink C, Smit E, Diviani N, Van Weert JC. Health literacy and online health information processing: unraveling the underlying mechanisms. J Health Commun 2016;21(sup2):109-120. [doi: 10.1080/10810730.2016.1193920] [Medline: 27668318]

102. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009 Oct;62(10):1013-1020. [doi: 10.1016/j.jclinepi.2008.10.009] [Medline: 19230606]

## Abbreviations

**CDC:** Centers for Disease Control and Prevention
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

---

*Edited by R Kukafka; submitted 06.04.22; peer-reviewed by C Lorini, N Hu; comments to author 31.05.22; revised version received 13.06.22; accepted 19.07.22; published 13.09.22.*

*Please cite as:*
*Ceretti E, Covolo L, Cappellini F, Nanni A, Sorosina S, Beatini A, Taranto M, Gasparini A, De Castro P, Brusaferro S, Gelatti U*
*Evaluating the Effectiveness of Internet-Based Communication for Public Health: Systematic Review*
*J Med Internet Res 2022;24(9):e38541*
*URL: https://www.jmir.org/2022/9/e38541*
*doi:10.2196/38541*
*PMID:36098994*

Review

# Consumers' Willingness to Pay for eHealth and Its Influencing Factors: Systematic Review and Meta-analysis

Zhenzhen Xie[1], PhD; Jiayin Chen[1], MSc; Calvin Kalun Or[1], PhD

Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong, China (Hong Kong)

**Corresponding Author:**
Calvin Kalun Or, PhD
Department of Industrial and Manufacturing Systems Engineering
The University of Hong Kong
Room 8-7, 8/f, Haking Wong Building
The University of Hong Kong, Pokfulam Road
Hong Kong
China (Hong Kong)
Phone: 852 3917 2587
Email: klor@hku.hk

## Abstract

**Background:** Despite the great potential of eHealth, substantial costs are involved in its implementation, and it is essential to know whether these costs can be justified by its benefits. Such needs have led to an increased interest in measuring the benefits of eHealth, especially using the willingness to pay (WTP) metric as an accurate proxy for consumers' perceived benefits of eHealth. This offered us an opportunity to systematically review and synthesize evidence from the literature to better understand WTP for eHealth and its influencing factors.

**Objective:** This study aimed to provide a systematic review of WTP for eHealth and its influencing factors.

**Methods:** This study was performed and reported as per the Cochrane Collaboration and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. PubMed, CINAHL Plus, Cochrane Library, EconLit, and PsycINFO databases were searched from their inception to April 19, 2022. We conducted random-effects meta-analyses to calculate WTP values for eHealth (at 2021 US dollar rates) and meta-regression analyses to examine the factors affecting WTP.

**Results:** A total of 30 articles representing 35 studies were included in the review. We found that WTP for eHealth varied across studies; when expressed as a 1-time payment, it ranged from US \$0.88 to US \$191.84, and when expressed as a monthly payment, it ranged from US \$5.25 to US \$45.64. Meta-regression analyses showed that WTP for eHealth was negatively associated with the percentages of women ($\beta=-.76$; $P<.001$) and positively associated with the percentages of college-educated respondents ($\beta=.63$; $P<.001$) and a country's gross domestic product per capita (multiples of US \$1000; $\beta=.03$; $P<.001$). Compared with eHealth provided through websites, people reported a lower WTP for eHealth provided through asynchronous communication ($\beta=-1.43$; $P<.001$) and a higher WTP for eHealth provided through medical devices ($\beta=.66$; $P<.001$), health apps ($\beta=.25$; $P=.01$), and synchronous communication ($\beta=.58$; $P<.001$). As for the methods used to measure WTP, single-bounded dichotomous choice ($\beta=2.13$; $P<.001$), double-bounded dichotomous choice ($\beta=2.20$; $P<.001$), and payment scale ($\beta=1.11$; $P<.001$) were shown to obtain higher WTP values than the open-ended format. Compared with ex ante evaluations, ex post evaluations were shown to obtain lower WTP values ($\beta=-.37$; $P<.001$).

**Conclusions:** WTP for eHealth varied significantly depending on the study population, modality used to provide eHealth, and methods used to measure it. WTP for eHealth was lower among certain population segments, suggesting that these segments may be at a disadvantage in terms of accessing and benefiting from eHealth. We also identified the modalities of eHealth that were highly valued by consumers and offered suggestions for the design of eHealth interventions. In addition, we found that different methods of measuring WTP led to significantly different WTP estimates, highlighting the need to undertake further methodological explorations of approaches to elicit WTP values.

***(J Med Internet Res 2022;24(9):e25959)*** doi:10.2196/25959

**KEYWORDS**

systematic review; meta-analysis; willingness to pay; eHealth; contingent valuation; discrete choice experiment; mobile phone

# Introduction

Advances in broadband technology and the Internet of Things have enabled the broad implementation of eHealth—the provision or acquisition of health information or services through electronic processes [1-7]. In recent years, a broad spectrum of eHealth interventions using various modalities has been developed and examined in health care research. Examples include websites, diagnostic and monitoring devices, smartphone apps, virtual reality systems, telephone and video calls, and electronic messages that provide health information or services [8-10]. Researchers have implemented these eHealth interventions into a range of health care activities, including teleconsultation [11,12], remote patient monitoring [13], self-management of diseases [14-16], disease rehabilitation [17], and disease prevention [18]. Promising results have emerged from these studies, which showed that eHealth interventions could facilitate the delivery of health care and improve patient outcomes [9-17]. It has also been shown that eHealth enables consumers to easily obtain information about health issues for decision-making, which could lead to more effective care, patient empowerment, and time savings [8,18-22].

Although eHealth is considered a promising complement to conventional health care systems, there are significant costs involved in its implementation arising from the purchase, development, and maintenance of hardware and software [23]. Therefore, when deciding to implement eHealth for personal use or public health, decision-makers need solid evidence that the costs of eHealth can be justified by its benefits [24]. This requires the quantification and measurement of the benefits of eHealth, which can then be aggregated with the costs of eHealth to understand its cost-effectiveness [25].

To measure the benefits of eHealth, willingness to pay (WTP) is a commonly used metric [26,27]. Welfare economics defines WTP as the maximum amount of money an individual is willing to pay for 1 unit of a good or service; it is an accurate proxy for the welfare (benefits) derived from that good or service [28-30]. A major advantage of the WTP approach is that it summarizes the benefits in monetary terms, making it comparable with the costs for use in cost-benefit analyses [26,31]. Another advantage is that WTP illustrates the perceived benefits from the perspective of consumers, which can be further analyzed to represent consumer preferences [32,33]. Therefore, the WTP approach is suitable for measuring the benefits of eHealth, as it can generate findings for the effective implementation of eHealth and provide insights into designing better eHealth technology and services.

Many studies [34-36] have examined consumers’ WTP for eHealth using either of the 2 mainstream methods. The first is contingent valuation, a survey-based method in which people are asked to indicate the maximum price they are willing to pay for eHealth (eg, services) or associated eHealth technology. The second is the discrete choice experiment, sometimes referred to as conjoint analysis, which involves asking people to state their preference for hypothetical alternatives that describe eHealth or eHealth technology. Regardless of the methods used, these studies have provided insights into consumers’ perceived eHealth benefits and the factors affecting these perceptions. If we synthesize and analyze these studies, we can obtain practical implications for the design, development, and implementation of eHealth and suggestions for future research. Thus, we systematically reviewed previous studies on consumers’ WTP for eHealth and synthesized their findings through a meta-analysis to understand consumers’ WTP for eHealth and examine its influencing factors. To the best of our knowledge, this is the first study of its kind.

# Methods

## Overview

This review was performed and reported according to the Cochrane Collaboration [37] and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1 provides the checklist) [38]. A total of 2 researchers (ZX and JC) independently screened the titles and abstracts of the articles identified in the literature search for eligibility, reviewed the full texts of potentially eligible articles for final inclusion in the review, extracted data from the final sample, critically appraised their methodological quality, and assessed the quality of the evidence. All disagreements between them were resolved through a consensus-based discussion.

## Search Strategy

We searched PubMed, CINAHL Plus, Cochrane Library, EconLit, and PsycINFO databases from their inception to April 19, 2022, to obtain a preliminary list of relevant studies. A search strategy was developed based on the following concepts combined using “AND”: WTP, money, and eHealth. For each concept, a set of keywords and their synonyms and variations were developed and combined in the search strategy using “OR.” The following search terms were developed: (“willingness to pay” *OR* “WTP” *OR* “valuation” *OR* “preference”) *AND* (“cost” *OR* “price” *OR* “expense” *OR* “money”) *AND* (“eHealth” *OR* “electronic health” *OR* “digital health” *OR* “mHealth” *OR* “mobile” *OR* “web” *OR* “Internet” *OR* “online” *OR* “tele*” *OR* “medical informatics” *OR* “medical information systems”). These search terms were used to search for titles and abstracts in all the selected databases, with no filters or limits placed on the search.

## Eligibility Criteria and Study Selection

We included all studies that (1) recruited participants who were consumers of eHealth, (2) measured and reported participants’ WTP for eHealth or eHealth technology, and (3) were published in a peer-reviewed English-language journal. Studies were excluded if they examined WTP from a public payer’s perspective (eg, WTP for public health programs through taxation) or a caregiver’s perspective (eg, parents’ WTP for their children). We also excluded reviews, case studies, poster presentations, and conference presentations but examined their references to identify additional relevant articles for inclusion. We also manually searched the reference lists of studies in the final sample for additional relevant articles.

### Data Extraction and Management

We extracted the following data from each study: country where the study was conducted, year in which it was conducted, sample size, sample characteristics, modality used to provide eHealth, details of the eHealth examined, WTP, method used to measure WTP, and WTP factors examined. Regarding the methods used to measure WTP, the extracted information included the formats of the questions posed to the study participants (eg, open-ended questions, dichotomous choice, and bidding games), whether the participants had used eHealth at the time of evaluation (ex post or ex ante), and how zero responses were dealt with (all zero responses excluded, all zero responses included, or protest zero responses excluded). We contacted the authors for clarification and verification of cases where relevant data were missing or incomplete.

### Critical Appraisal of Methodological Quality

The included studies were critically appraised for methodological quality using 17 criteria based on the Hoy risk of bias assessment tool [39] and a set of criteria specific for assessing WTP studies (Multimedia Appendix 2) [29,40].

### Data Analysis

#### *Descriptive Statistics and Narrative Synthesis of the Studies in the Final Sample*

Descriptive statistics were used to summarize the characteristics of the included studies. Narrative synthesis was used to synthesize the WTP findings for eHealth in the studies, for which the means, SDs, 95% CIs, medians, IQRs, and ranges were reported. All WTP values were calculated at 2021 US dollar rates to facilitate quantitative synthesis and comparison. First, the WTP values in other currencies were converted to US dollars based on the purchasing power parity (PPP) exchange rate of the year in which the study was conducted, and then they were converted to 2021 US dollar values using gross domestic product (GDP) deflators. The PPP exchange rate and GDP deflator data were obtained from the International Monetary Fund’s World Economic Outlook database [41]. For studies that did not report the year in which they were conducted, we used the year preceding the publication year of the articles for currency conversion.

#### *Random-Effects Meta-analyses to Measure WTP*

We performed random-effects meta-analyses to estimate the overall WTP value for eHealth and the WTP value for eHealth by different subgroups (ie, modalities used to provide eHealth and the region where the study was conducted) [42]. The WTP values were log-transformed to reduce skewness [43]. In the meta-analysis, the weight of each study was the inverse of the WTP variance. For studies that did not report variance (or SD), we obtained an estimate using (1) SE and sample size, (2) 95% CIs and sample size, (3) IQRs, or (4) range and sample size [37,44]. The $I^2$ test was used to measure heterogeneity in the synthesized studies [45], and the Egger test was used to assess the possibility of publication bias [46].

#### *Meta-regression Analyses to Examine the Factors Affecting WTP*

Univariate meta-regressions were conducted to examine whether WTP for eHealth was influenced by explanatory variables, including gender, age, and education level of the study sample; per capita GDP of the country where the study was conducted and the year in which it was conducted; the modality used to provide eHealth (ie, websites, medical devices, health apps, asynchronous communication, and synchronous communication); the format of the WTP questions (ie, open-ended, single-bounded dichotomous choice, double-bounded dichotomous choice, or payment scale); whether the participants of the study had used eHealth at the time of evaluation (ex post vs ex ante); and whether zero responses were excluded from the analysis of WTP values. A mixed-effects log-linear regression model was used, where the payment horizon (1-time or monthly payment) was modeled as a random effect and the explanatory variable was modeled as a fixed effect. We also narratively synthesized the WTP factors for eHealth examined in the included studies. All statistical analyses were performed in R (version 4.0.2, R Foundation for Statistical Computing) software using the *metafor* package.

### Assessment of the Quality of Evidence

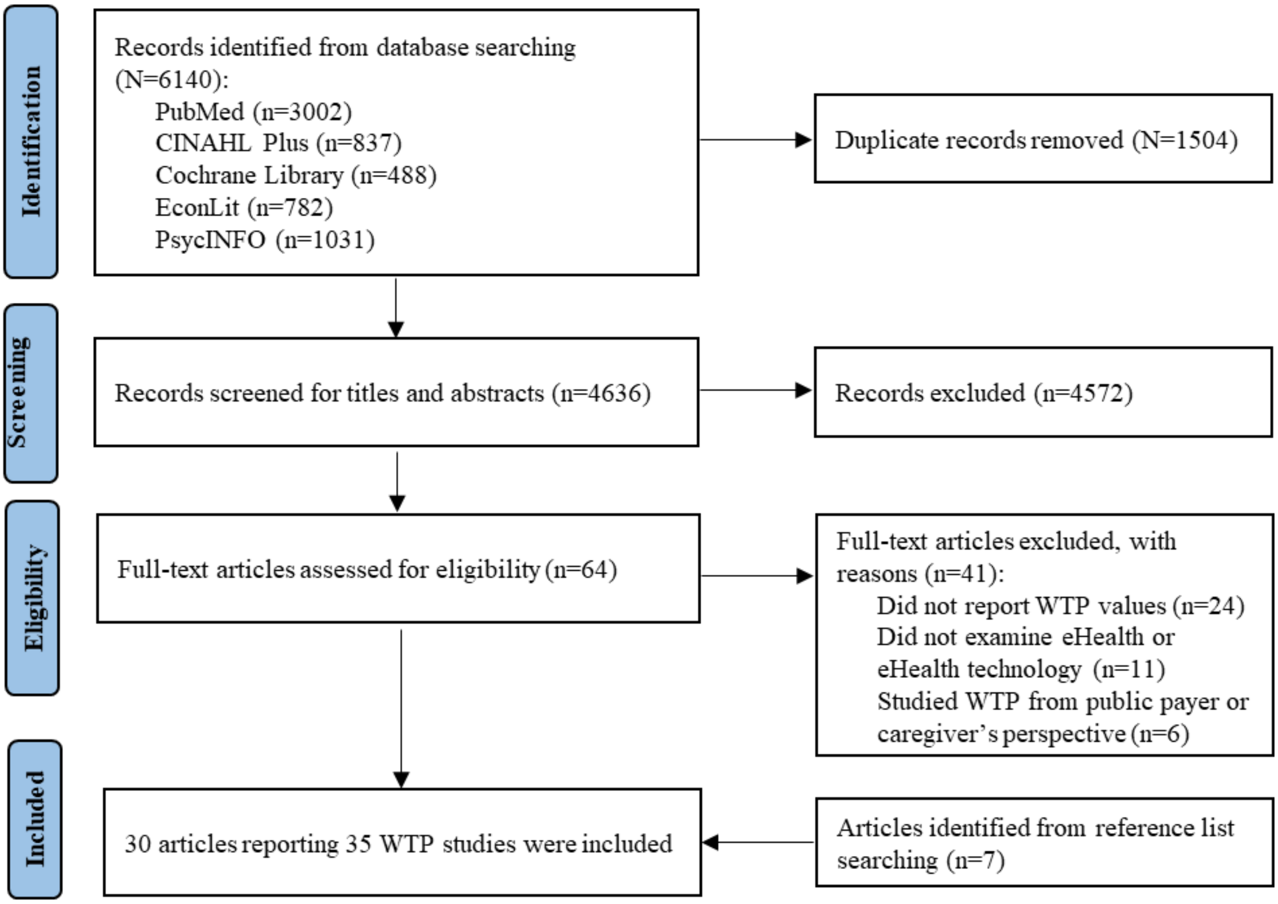
The quality of evidence of the meta-analysis results was assessed using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) framework [47]. We adopted the framework for rating the relative importance of outcomes (eg, values, preferences, and outcome importance) [48,49], which was more suitable for rating cross-sectional WTP surveys and discrete choice experiments than previous GRADE guidelines that focused on the effects of interventions. For each WTP outcome, the quality of evidence started from “high” and was downgraded by 1 level for every serious issue identified in the domains of risk of bias, inconsistency, indirectness, imprecision, and publication bias. The risk of bias domain was assessed by inspecting the potential bias in participant selection, measurement instruments, data collection, and data analysis. The inconsistency domain was assessed using $I^2$ values, and the GRADE quality was downgraded when $I^2 \geq 50\%$. The indirectness domain was assessed using the indirectness of the population, outcomes, options, and methodologies used to elicit the values of the outcomes. The imprecision domain was assessed using the width of the CIs of the estimates and sample size. The publication bias domain was assessed using the Egger test, and GRADE quality was downgraded for statistically significant findings ($P<.05$) on this test.

## Results

### Literature Search and Selection Process

Figure 1 shows the literature search and selection process. The search yielded 6140 articles, of which 30 (0.49%) articles representing 35 WTP studies were identified as eligible and included in the final review.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the study selection process. WTP: willingness to pay.



## Study Characteristics

The characteristics of the studies included in this review are presented in Table 1. Appraisals of the methodological quality are presented in Multimedia Appendix 2.

**Table 1.** Summary of the characteristics of the final studies (N=35).

| Characteristics | Values, n (%) |
|---|---|
| **Study location** | |
| Africa | 2 (6) |
| Asia | 8 (23) |
| Europe | 13 (37) |
| North America | 9 (26) |
| Oceania | 3 (9) |
| **Year of publication** | |
| 2003-2010 | 8 (23) |
| 2011-2015 | 8 (23) |
| 2016-2021 | 19 (54) |
| **Modality used to provide eHealth** | |
| Websites | 5 (14) |
| Medical devices | 8 (23) |
| Health apps | 5 (14) |
| Asynchronous communication (eg, SMS text messaging or email) | 8 (23) |
| Synchronous communication (eg, telephone call or video call) | 7 (20) |
| Not specified | 2 (6) |
| **Method used to measure willingness to pay** | |
| **Contingent valuation** | 26 (74) |
| Open-ended questions | 13 (37) |
| Single-bounded dichotomous choice questions | 1 (3) |
| Double-bounded dichotomous choice questions | 4 (11) |
| Payment scale questions | 2 (6) |
| Bidding games | 2 (6) |
| Single-bounded dichotomous choice+payment scale questions | 1 (3) |
| Not reported | 3 (9) |
| Discrete choice experiment | 9 (26) |

## WTP for eHealth: Narrative Synthesis

Table 2-Table 4 present the details of 74% (26/35) of studies that used contingent valuations and 26% (9/35) of studies that used discrete choice experiments.

**Table 2.** Details of the 26 contingent valuation studies included in the final sample.

| Study | Country (year of study) | Population and sample size (N) | Age (years) | Women (%) | eHealth details | Measurement of WTP[a] (format, ex ante or ex post, and zeros) | WTP (PPP[b], and 2021 US dollar value) |
|---|---|---|---|---|---|---|---|
| **Contingent valuation studies that reported WTP as a 1-time payment (n=17)** | | | | | | | |
| Adedokun et al [50] | Nigeria (2011) | Patients at a family medicine unit (389) | Mean 42.1 | 54 | An SMS text messaging–based appointment scheduling service: patients sent an SMS text message to book a clinic appointment and received a confirmation SMS text message and another SMS text message reminding them of the appointment | Open-ended; ex ante; all zeros excluded | Mean 2.81 (SD 3.88), range 0.06-38.26 |
| Belkora et al [51] | United States (2007-2010) | Patients with breast cancer (34) | Mean 59 | 100 | A telephone consultation planning service: before a clinical visit, a community health worker called the patient to check if they had any medical questions and then sent the list of questions to the patient's physician | Double-bounded dichotomous choice; ex post; all zeros included | Mean 191.84 (SD 242.91) |
| Bergmo and Wangberg (1) [52] | Norway (2003) | Patients at a primary clinic (52) | Mean 38 | 70 | An internet-based messaging system that enabled patients to communicate with their health care providers by sending messages using a web browser | Open-ended; ex ante; protest zeros excluded | Mean 10.94 (95% CI 8.91-13.17); median 10.14 (IQR 5.07-20.26) |
| Bergmo and Wangberg (2) [52] | Norway (2003) | Patients at a primary clinic (38) | Mean 37 | 61 | Same as Bergmo and Wangberg (1) [52] | Open-ended; ex post; protest zeros excluded | Mean 7.30 (95% CI 5.47-8.91); median 7.09 (IQR 2.03-10.14) |
| Brandling-Bennet et al [53] | Cambodia (2003) | Patients at a clinic (49) | Mean 39 | 61 | A telemedicine service: local nurses recorded the medical history and conducted physical examinations of patients and sent this information to physicians at a remote place via email; the physicians would then reply with the treatment or referral decisions; the local nurses would execute the recommendations | Not reported; ex post; all zeros excluded | Median 0.90, range 0-72.53 |

| Study | Country (year of study) | Population and sample size (N) | Age (years) | Women (%) | eHealth details | Measurement of WTP[a] (format, ex ante or ex post, and zeros) | WTP (PPP[b], and 2021 US dollar value) |
|---|---|---|---|---|---|---|---|
| Fawsitt et al (1) [54] | Ireland (2015) | Women in antenatal clinics (20) | Mean and median not reported | 100 | A mobile app that provided information about cesarean section and surgical site infections: users recorded symptoms, temperature, heart rate, and pain level based on which the app would provide health advice (eg, check body temperature or contact a general practitioner) | Open-ended; ex ante; all zeros included | Mean 30.96 (SD 58.28); median 13.98 |
| Fawsitt et al (2) [54] | Ireland (2015) | Women in antenatal clinics (116) | Mean and median not reported | 100 | A mobile app that provided information about cesarean section and surgical site infections: users recorded symptoms, temperature, heart rate, and pain level, which would be checked daily by a midwife in the maternity hospital who would provide health advice to the user | Open-ended; ex ante; all zeros included | Mean 36.38 (SD 51.46); median 13.98 |
| Fawsitt et al (3) [54] | Ireland (2015) | Women in antenatal clinics (44) | Mean and median not reported | 100 | A telephone call–based helpline service: users called a midwife in the maternity hospital, who would provide health advice and instructions | Open-ended; ex ante; all zeros included | Mean 32.76 (SD 47.73); median 13.98 |
| Kaga et al [55] | Japan (2016) | General population (305) | Mean and median not reported | 37 | An internet-based telecare service for older adults, which connected the television at users' homes to the internet: health care information was displayed on the television; if the television was not used for 3 days, a telephone call would be made to the user, and if they did not answer the call, neighborhood associations and civil servant committees would visit them to ensure that they were fine | Double-bounded dichotomous choice; ex ante; all zeros included | Mean 8.58; median 4.57 |
| Ngan et al [56] | Vietnam (2017) | Smokers who intended to quit (433) | Mean 33 | 0.8 | An SMS text messaging–based smoking cessation service: SMS text messages with relevant health information, suggestions for controlling and preventing cravings, and encouragement were sent to users 2 to 4 times a day for 6 weeks | Single-bounded dichotomous choice; ex ante; all zeros included | Mean 59.99 (95% CI 46.92-73.07) |

| Study | Country (year of study) | Population and sample size (N) | Age (years) | Women (%) | eHealth details | Measurement of WTP[a] (format, ex ante or ex post, and zeros) | WTP (PPP[b], and 2021 US dollar value) |
|---|---|---|---|---|---|---|---|
| Raghu et al (1) [57] | United States (2013-2014) | Patients waiting for general consultation (214) | Mean and median not reported | Not reported | A teledermoscopy service: a clinician at a health center used a smartphone (with a Canfield Dermscopefield) to capture images of skin lesions and send them to a dermatologist, who then wrote a medical note and sent it to the clinician | Double-bounded dichotomous choice; ex ante; all zeros included | Mean 63.12 (SD 44.66); median 55.77 |
| Raghu et al (2) [57] | United States (2013-2014) | Patients with skin lesions (41) | Mean and median not reported | Not reported | Same as Raghu et al (1) [57] | Double-bounded dichotomous choice; ex ante; all zeros included | Mean 59.81 (SD 30.33); median 54.83 |
| Ramchandran et al [58] | United States (2017) | Patients with diabetes (23) | Mean 56 | 52 | A teleophthalmology service: a technician or nurse used a nonmydriatic fundus camera to take photos of the patient's eye and send them to an ophthalmologist, who then replied with a diagnosis and recommended follow-up care | Payment scale; ex ante; all zeros included | Mean 29.96 (SD 8.53) |
| Rochat et al [59] | Switzerland (2014) | People visiting a travel clinic (162) | Mean and median not reported | 53 | A telemedicine service for travelers providing pretravel information; medical advice for upcoming trips; and health advice when the traveler was abroad through telephone calls, video calls, or emails | Not reported; ex ante; all zeros excluded | Median 57.10 (IQR 34.26-57.10) |
| Ruby et al [60] | United States (2008) | Adolescents with persistent subthreshold depression (34) | Mean 17 | 57 | An internet-based depression prevention intervention for adolescents: 14 modules for depression prevention were provided through a website | Not reported; ex post; all zeros included | Median 50.15 (IQR 19.59-62.68); range 0-626.84 |
| Shariful Islam et al [61] | Bangladesh (2013-2014) | Patients with type 2 diabetes (352) | Mean 50 | 56 | An SMS text message–based health service for patients with type 2 diabetes, which provided medication reminders and relevant health information (eg, diabetes complications and recommended diet and physical activities) through SMS text messages | Open-ended; ex ante; all zeros included | Median 0.88 (IQR 1.99) |
| Stahl et al [62] | United States (2007-2008) | Patients visiting a primary care physician (101) | Mean 46 | 60 | An internet-based primary care service: a primary care physician took the patient's medical history, conducted a visual inspection, decided on treatment, and arranged follow-up care through videoconferencing | Payment scale; ex post; all zeros included | Mean 25.71 (SD 15.88)[c] |

| Study | Country (year of study) | Population and sample size (N) | Age (years) | Women (%) | eHealth details | Measurement of WTP[a] (format, ex ante or ex post, and zeros) | WTP (PPP[b], and 2021 US dollar value) |
|---|---|---|---|---|---|---|---|
| **Contingent valuation studies that reported WTP as monthly payments: WTP per month (n=9)** | | | | | | | |
| Cocosila et al [63] | Canada (2006-2007) | General population (51) | Median 21 | 57 | An SMS text message–based health reminder service: users received SMS text messages reminding them to take vitamin C pills | Open-ended; ex post; no zero responses | Median 5.25; range 0.52-31.47 |
| Contreras-Somoza et al [64] | Spain, Serbia, Netherlands, France, Israel, Italy, or Slovenia (not reported) | Patients aged >60 years with mild cognitive impairment (30) | Mean 73.3 | 60 | An internet-based information and communication technology platform (ehcoBUTLER system) for older people: the platform hosted several social and health apps to support the daily activities of older people and improve their health, quality of life, and independence | Not reported; ex ante; all zeros excluded | Median 14.64 |
| Jacobs et al [65] | Belgium (2009) | General population (135) | Mean 41 | 34 | A cardiovascular disease prevention program with internet-based components: the program comprised cardiovascular risk assessment, communication, follow-up care, a website providing health information on cardiovascular disease, advice on physical activity and diet, guidelines for behavioral changes, and individual coaching by a health psychologist | Single-bounded dichotomous choice+payment scale; ex post; all zeros included | Mean 13.41 (SD 14.42); median 5.64 |
| Rasche et al [66] | Germany (2017) | General population (96) | Mean 63.8 | 51 | A mobile app for fall prevention: the app had features such as detecting the risk of falling, recommendations for reducing this risk, storing other health-related data, and providing advice on how to prevent and respond to a fall | Open-ended; ex ante; all zeros included | Median 7.41 (IQR 14.83); range 0-118.61 |
| Somers et al (1) [34] | United Kingdom (2015) | General population (1697) | Mean 47 | 51 | A mobile app for improving well-being outcomes: the app had features such as calling and messaging friends or families or local health care providers, setting health goals, tracking health status, sharing health data, and receiving information about the local community | Open-ended; ex ante; all zeros included | Mean 24.31; median 7.46; range 0-1344.36 |

| Study | Country (year of study) | Population and sample size (N) | Age (years) | Women (%) | eHealth details | Measurement of WTP[a] (format, ex ante or ex post, and zeros) | WTP (PPP[b], and 2021 US dollar value) |
|---|---|---|---|---|---|---|---|
| Somers et al (2) [34] | United Kingdom (2015) | General population (305) | Mean 48 | 72 | Same as Somers et al (1) [34]. | Open-ended; ex ante; ell zeros included | Mean 20.13; median 7.46; range 0-896.62 |
| Tran et al [67] | Vietnam (2012) | Patients with HIV or AIDS (1016) | Mean 35.4 | 36 | A mobile phone–based medication reminder service for patients with HIV: SMS text messages, telephone calls, or automated voice calls were used to remind patients to take their medication on time | Not reported; ex ante; all zeros included | Mean 8.42 |
| Tsuji et al [68] | Japan (not reported) | General population (291) | Mean and median not reported | Not reported | A telehealth system for older people: health-related data such as blood pressure, oxygen saturation, heart rhythm, electrical activity, and heart rates were measured at the user's home and sent to a remote clinic where nurses studied them and reported any unusual symptoms to the user and physicians; monthly health reports were created and sent to users | Bidding game; ex post; all zeros included | Mean 45.64 |
| Tsuji et al [69] | Japan (not reported) | General population (145) | Mean 74 | 74 | Same as Tsuji et al [68] | Bidding game; ex ante; all zeros included | Mean 29.68 |

[a] WTP: willingness to pay.

[b] PPP: purchasing power parity.

[c] The WTP values were obtained by combining the WTP values for subgroups, as reported in the articles.

**Table 3.** Demographic and eHealth details of the 9 discrete choice experiment studies included in the final sample.

| Study | Country (year of study) | Population | Sample size, N | Age (years) | Women (%) | eHealth details |
|---|---|---|---|---|---|---|
| **Discrete choice experiment studies that reported WTP[a] as a 1-time payment (n=6)** | | | | | | |
| Buchanan et al [35] | United Kingdom (2018) | General population | 734 | Mean 47 | 51 | Web-based consultation with a primary care physician |
| Park et al [70] | South Korea (2009-2010) | Patients in endocrinology and metabolism clinics | 118 | Mean 57 | 58 | A telemedicine service for patients with diabetes |
| Snoswell et al [71] | Australia (not reported) | General population | 113 | Mean 40 | 74 | A mobile teledermoscopy service for skin cancer screening: users used a dermoscopic smartphone attachment and app to take photos and send them to a dermatologist, along with relevant clinical information |
| Snoswell et al [36] | Australia (2019) | Patients who had a video consultation in the previous year | 62 | Mean and median not reported | 62.9 | Web-based consultation with a specialist physician through videoconferencing |
| Spinks et al [72] | Australia (not reported) | People aged 50 to 64 years at high risk of melanoma | 35 | Mean and median not reported | 54 | A teledermoscopy service for skin cancer screening: using a dermatoscope to take photos which were sent to a dermatologist for diagnosis |
| van der Pol and McKenzie [73] | United Kingdom (not reported) | General population | 90 | Mean and median not reported | 62 | A telemedicine service for ear, nose, and throat examination: patients sent endoscopic images to and videoconferenced with a specialist |
| **Discrete choice experiment studies that reported WTP as monthly payments (n=3)** | | | | | | |
| Ahn et al [74] | South Korea (2011) | General population | 400 | Mean 44 | 51 | A telemedicine service system that measured vital signs of users and transmitted patient data to care providers |
| Chang et al [75] | United States (2009-2010) | General population | 6271 | Mean and median not reported | 52 | A web-based health service that provided remote diagnosis, treatment, monitoring, and consultation |
| Deal et al [76] | Canada (not reported) | Patients with cardiovascular disease | 74 | Mean 68.9 | 50 | A web-based system that tracked and displayed patients' details on 15 outcomes related to cardiovascular disease risk, the target value of these outcomes for better control of their condition, the last time the outcome was checked, and brief advice for patients and clinicians |

[a] WTP: willingness to pay.

**Table 4.** WTP[a] details of the 9 discrete choice experiment studies included in the final sample.

| Study, attribute (reference level), and desired level or levels of the attribute | Marginal WTP (PPP[b], 2021 US dollar value) |
|---|---|
| **Discrete choice experiment studies that reported WTP as a 1-time payment (n=6)** | |
| **Buchanan et al [35]** | |
| **How similar was your consultation to a traditional “face-to-face” appointment (the same)** | |
| Video consultation | –7.02 |
| Symptoms submitted via an electronic form | –15.40 |
| **How long did you have to wait for a consultation** | |
| Reduced by 1 hour | 0.22 |
| **Reputation of the GP[c] (2 stars)** | |
| 5 stars | 13.65 |
| **Collecting antibiotics (taking a paper prescription to a pharmacy located in the same building as the local medical center)** | |
| Prescription emailed to a pharmacy in another building as the local medical center | –11.38 |
| **Form of consultation (at local medical centers)** | |
| Via the internet (–10.83) | –17.09 |
| **Park et al [70]** | |
| **Service platform (the internet)** | |
| Mobile phone | 22.72 |
| **Service providers (small- and medium-sized hospitals and clinics)** | |
| Large general hospitals | 21.64 |
| **Service scope (glucose management only)** | |
| Comprehensive diabetes care | 24.23 |
| **Personalization of consultation (absent)** | |
| Present | 11.87 |
| **24-hour service accessibility (absent)** | |
| Present | 10.27 |
| **Reply time (within 3 days)** | |
| Within 1 day | 8.45 |
| **Assurance of service (low assurance)** | |
| High assurance | 18.61 |
| **System failure (system down 1%-5%)** | |
| System down <1% | 12.68 |
| **Confidentiality (1%-5% confidentiality breaches)** | |
| <1% confidentiality breaches | 8.78 |
| **Snoswell et al [71]** | |
| **Method of screening (by a GP)** | |
| Mobile teledermoscopy | 0.88 |
| **Time away from usual activities (>4 hours)** | |
| 3-4 hours | 6.11 |
| 1-2 hours | 53.75 |
| **Chances of detecting a melanoma if one is present (65%–75%)** | |
| 85%-95% | 54.37 |
| ≥95% | 87.73 |

| Study, attribute (reference level), and desired level or levels of the attribute | Marginal WTP (PPP[b], 2021 US dollar value) |
|---|---|
| **Wait time for results (3 days)** | |
| <4 hours | 4.92 |
| **Person reviewing the results (GP)** | |
| Dermatologist | 32.21 |
| **Number of moles removed to find 1 melanoma (5)** | |
| 3 | 31.51 |
| **Snoswell et al [36]** | |
| **Type and mode of consultation (local in-person consultation with a generalist physician at a GP clinic or small hospital)** | |
| In-person consultation with a specialist physician at a large metropolitan hospital | 9.88 |
| Videoconference with a specialist physician from a local GP clinic or small hospital | 91.33 |
| Videoconference with a specialist from home | 33.53 |
| **Time away from home, office, or usual activities, including travel (half a day)** | |
| 1 full day | –11.80 |
| ≥2 full days | –113.66 |
| **Perceived benefit from the consultation (limited)** | |
| Partial benefit | 53.86 |
| Benefit | 111.28 |
| **Consulted or not (attending a consultation)** | |
| No consultation chosen | –175.14 |
| **Spinks et al [72]** | |
| WTP for teledermoscopy service, in addition to skin self-examination, GP screening, and clinic skin cancer screening | 84.38 |
| **van der Pol and McKenzie [73]** | |
| **Type of clinic** | |
| Telemedicine | 773.31 |
| Face-to-face | 873.45 |
| **Driving time (up to 30 minutes)** | |
| 30-60 minutes | –57.49 |
| 60-90 minutes | –74.18 |
| 2-4 hours | –155.77 |
| **Wait time** | |
| Each additional week | –27.82 |
| **Discrete choice experiment studies that reported WTP as a monthly payment: WTP per month (n=3)** | |
| **Ahn et al [74]** | |
| **Device type (smartphone)** | |
| Smart home | 138.29 |
| Wearable device | 632.49 |
| **Service type (management of oxygen saturation level)** | |
| Blood glucose | 30.27 |
| Blood pressure | –56.35 |
| **Service tailoring (absent)** | |
| Present | 82.76 |

| Study, attribute (reference level), and desired level or levels of the attribute | Marginal WTP (PPP[b], 2021 US dollar value) |
|---|---|
| **Reply time (usual)** | |
| 1-hour reduction | 3.57 |
| **Chang et al [75]** | |
| Per household | 5.40[d] |
| **Deal et al [76]** | |
| **Speed of adding new information to the system (2 weeks)** | |
| 1 week | 5.70 |
| 48 hours | 7.60 |
| Overnight | 2.85 |
| 1 hour | 0 |
| **Individual patient tracker values displayed (most recent values only)** | |
| 2 most recent | 8.55 |
| 12-month history | 13.31 |
| 5-year history | 8.55 |
| Complete history | –5.70 |
| **Nurse coordinator tasks or duties (no nurse coordinator)** | |
| Basic duties[e] | 16.16 |
| Basic duties and input data | 20.91 |
| Basic duties and information sessions | 17.11 |
| Basic duties, phone, and email | 33.27 |
| Basic duties and reminders | 19.96 |
| **Frequency of contacting nurse coordinator (no contact)** | |
| 1 day per month | 6.65 |
| 2 days per month | 10.45 |
| 1 day per week | 5.70 |
| 2 days per week | 10.91 |
| 5 days per week | 1.90 |
| **Number of visits to a physician per year (1)** | |
| 2 | 19.01 |
| 3 | 25.66 |
| 4 | 27.56 |
| 6 | 7.60 |

[a] WTP: willingness to pay.

[b] PPP: purchasing power parity.

[c] GP: general practitioner.

[d] 95% CI 3.79-7.02.

[e] Basic duties of the nurse coordinator: assist the physician in using the tracker, keep tracker information updated, and ensure action is taken to address uncontrolled cardiovascular disease risks.

## WTP for eHealth: Meta-analysis

Approximately 60% (21/35) of studies reported sufficient data for inclusion in the meta-analysis. Among the 21 studies, 16 (76%) reported that WTP was measured as a 1-time payment, whereas 5 (24%) reported that it was measured as monthly payments. Table 5 presents the mean WTP for eHealth obtained through the meta-analysis.

**Table 5.** Overall WTP[a] for eHealth, WTP by the modality used to provide eHealth, and WTP by the region where the study was conducted (N=21).

| Variables | Studies, n (%) | Sample size | WTP (PPP[b], 2021 US dollars), mean (95% CI) | $I^2$ (%) | Egger test | | GRADE[c] quality of evidence |
|---|---|---|---|---|---|---|---|
| | | | | | *z* score | *P* value | |
| **Studies that measured WTP as a 1-time payment (n=16) [50-62]** | | | | | | | |
| Overall WTP | 16 (76) | 2102 | 25.00 (12.79-48.87) | 99.69 | 0.72 | .47 | Low |
| **Modality used to provide eHealth** | | | | | | | |
| Websites | 1 (5) | 34 | 111.46 (84.55-146.92) | N/A[d] | N/A | N/A | Very low |
| Medical devices | 3 (14) | 278 | 48.34 (30.17-77.44) | 97.86 | 0.10 | .92 | Low |
| Health apps | 2 (10) | 136 | 35.86 (28.05-45.85) | N/A | N/A | N/A | Low |
| Asynchronous communication (eg, SMS text messages and email) | 6 (28) | 1313 | 7.76 (2.39-25.21) | 99.55 | 1.67 | .10 | Low |
| Synchronous communication (eg, telephone and video call) | 4 (19) | 341 | 52.59 (22.15-124.90) | 99.23 | 0.74 | .46 | Very low |
| **Region** | | | | | | | |
| North America | 6 (28) | 447 | 61.92 (33.94-112.97) | 99.01 | 3.30 | .001 | Low |
| Europe | 6 (28) | 432 | 22.65 (12.05-42.60) | 98.16 | 0.24 | .81 | Moderate |
| Asia | 3 (14) | 834 | 9.93 (0.84-117.06) | 99.76 | 1.3 | .19 | Low |
| Africa | 1 (5) | 389 | 2.81 (2.45-3.22) | N/A | N/A | N/A | Low |
| **Studies that measured WTP as monthly payments: WTP per month (n=5) [34,63,65,66]** | | | | | | | |
| Overall WTP | 5 (24) | 2284 | 18.53 (11.81-29.08) | 94.71 | 0.24 | .81 | Moderate |
| **Modality used to provide eHealth** | | | | | | | |
| Websites | 1 (5) | 135 | 13.41 (11.19-16.08) | N/A | N/A | N/A | Moderate |
| Health apps | 3 (14) | 2098 | 28.89 (21.71-38.44) | 44.49 | −1.73 | .08 | Moderate |
| Asynchronous communication | 1 (5) | 51 | 10.62 (8.89-12.68) | N/A | N/A | N/A | Low |
| **Region** | | | | | | | |
| North America | 1 (5) | 51 | 10.62 (8.89-12.68) | N/A | N/A | N/A | Low |
| Europe | 4 (19) | 2233 | 21.81 (13.91-34.20) | 91.27 | −0.12 | .91 | Moderate |

[a] WTP: willingness to pay.

[b] PPP: purchasing power parity.

[c] GRADE: Grading of Recommendations, Assessment, Development, and Evaluation.

[d] N/A: not applicable (as <3 experiments were analyzed).

Among the 16 studies that measured WTP as a 1-time payment, the mean WTP was US $25.00 (95% CI 12.79-48.87). The highest mean WTP was for eHealth provided through websites (US $114.46, 95% CI 84.55-146.92), followed by synchronous communication (US $52.59, 95% CI 22.15-124.90), medical devices (US $48.34, 95% CI 30.17-77.44), health apps (US $35.86, 95% CI 28.05-45.85), and asynchronous communication (US $7.76, 95% CI 2.39-25.21). In terms of region, the WTP value was the highest in North America (US $61.92, 95% CI 33.94-112.97), followed by Europe (US $22.65, 95% CI 12.05-42.60), Asia (US $9.93, 95% CI 0.84-117.06), and Africa (US $2.81, 95% CI 2.45-3.22).

Among the 5 studies that measured WTP as monthly payments, the mean WTP was US $18.53 (95% CI 11.81-29.08) per month. The highest mean WTP per month was for eHealth provided through health apps (US $28.89, 95% CI 21.71-38.44), followed by websites (US $13.41, 95% CI 11.19-16.08), and asynchronous communication (US $10.62, 95% CI 8.89-12.68). In terms of region, the mean WTP per month was US $21.81 (95% CI 13.91-34.20) in Europe and US $10.62 (95% CI 8.89-12.68) in North America.

### Factors Affecting WTP for eHealth: Meta-regression and Narrative Synthesis

Table 6 presents the results of the univariate log-linear meta-regression analyses of WTP-related factors for eHealth.

The results showed that higher percentages of women (β=−.76; *P*<.001) were associated with a lower mean WTP value for eHealth, and more people with a college education (β=.63; *P*<.001) were associated with a higher mean WTP value for eHealth. No significant evidence was found to support the association between age and WTP for eHealth (*P*=.57). A higher GDP per capita was found to be related to a higher WTP value for eHealth (β=.03; *P*<.001). Compared with eHealth provided through websites, the respondents had a lower WTP value for asynchronous communication (β=−1.43; *P*<.001) and a higher WTP value for medical devices (β=.66; *P*<.001) and synchronous communication (β=.58; *P*<.001). Studies eliciting WTP values using the single-bounded dichotomous choice format (β=2.13; *P*<.001), double-bounded dichotomous choice format (β=2.20; *P*<.001), payment scale format (β=1.11; *P*<.001), and unspecified formats (β=1.89; *P*<.001) had higher mean WTP values than those using open-ended formats. Ex post evaluations had lower WTP values (β=−.37; *P*<.001) than ex ante evaluations. However, there was no significant difference in WTP between studies that excluded protest zero responses or all zero responses and studies that included all zero responses in their analysis (*P*=.37).

Among the studies included in this review, 40% (14/35) examined WTP-related factors for eHealth, and their findings were narratively synthesized (Tables 7 and 8). The factors of interest included the characteristics of the eHealth technology or service and the study participants' sociodemographic characteristics, health conditions, current health care services, psychosocial characteristics, familiarity with information technology, and attitudes.

**Table 6.** Univariate log-linear meta-regression analyses of WTP[a]-related factors for eHealth.

| Explanatory variable | Outcome variable (mean WTP) | |
|---|---|---|
| | β (SE, 95% CI) | *P* value |
| Gender (women; %) | −.76 (0.14, −1.03 to −0.49) | <.001 |
| Age (years) | .002 (0.003, −0.004 to 0.01) | .57 |
| Education (completed college; %) | .63 (0.18, 0.29 to 0.98) | <.001 |
| GDP[b] per capita (US $) | .03 (0.001, 0.025 to 0.027) | <.001 |
| **Modality used to provide eHealth** | | |
| Websites | —[c] | — |
| Medical devices | .66 (0.08, 0.49 to 0.82) | <.001 |
| Health apps | .25 (0.1, 0.06 to 0.44) | .01 |
| Asynchronous communication (eg, SMS text messages and email) | −1.43 (0.09, −1.60 to −1.27) | <.001 |
| Synchronous communication (eg, telephone and video call) | .58 (0.08, 0.42 to 0.74) | <.001 |
| **WTP question format** | | |
| Open-ended | — | — |
| Single-bounded dichotomous choice | 2.13 (0.12, 1.90 to 2.36) | <.001 |
| Double-bounded dichotomous choice | 2.20 (0.06, 2.09 to 2.31) | <.001 |
| Payment scale | 1.11 (0.05, 1.01 to 1.21) | <.001 |
| Not reported | 1.89 (0.05, 1.80 to 1.98) | <.001 |
| Ex post vs ex ante | −.37 (0.04, −0.45 to −0.28) | <.001 |
| Protest zero or all zero responses excluded vs all zeros included | .02 (0.02, −0.02 to 0.05) | .37 |

[a] WTP: willingness to pay.

[b] GDP: gross domestic product.

[c] Not available because it was the reference level.

**Table 7.** WTP[a]-related factors for the examined eHealth in studies that reported WTP as a one-time payment.

| Factors | Adedokun et al [50] | Bergmo and Wangberg [52] | Kaga et al [55] | Ngan et al [56] | Raghu et al (1) [57] | Raghu et al (2) [57] | Shariful Islam et al [61] | Stahl et al [62] |
|---|---|---|---|---|---|---|---|---|
| **Characteristics of the eHealth or eHealth technology** | | | | | | | | |
| Favorable features | —[b] | — | — | — | — | — | — | Positive[c] |
| Technical quality | — | — | — | — | — | — | — | Not significant |
| Service convenience | — | — | — | — | Positive | Positive | — | — |
| Satisfaction with the service | — | — | — | — | Not significant | Not significant | — | — |
| Brand reputation | — | — | — | — | Positive | Positive | — | — |
| **Sociodemographic characteristics** | | | | | | | | |
| Gender (women) | Not significant | Not significant | Negative | — | — | — | Negative | Not significant |
| Age | Not significant | Positive | Not significant | Positive | — | — | — | Not significant |
| Education | Not significant | Not significant | — | Not significant | Not significant | Negative | Positive | — |
| Income | — | Not significant | Not significant | Positive | Not significant | Positive | Positive | — |
| Employment | Not significant | — | — | — | Not significant | Not significant | — | — |
| Occupation | Not significant | — | — | Not significant | — | — | — | — |
| Living alone | — | — | Not significant | — | — | — | — | — |
| Residential area | — | — | — | Not significant | — | — | — | — |
| International student | — | — | — | — | Negative | Not significant | — | — |
| **Health conditions** | | | | | | | | |
| Chronic conditions | — | Not significant | — | — | — | — | Not significant | — |
| Smoking status | — | — | — | Not significant | — | — | — | — |
| Attempts to quit smoking | — | — | — | Not significant | — | — | — | — |
| **Current health care services** | | | | | | | | |
| Number of visits to a physician | — | — | — | — | — | — | Not significant | — |
| Time taken and cost of travel to see a physician | — | — | — | — | — | — | — | Not significant |
| **Psychosocial characteristics** | | | | | | | | |
| Health anxiety | — | — | Not significant | — | — | — | — | — |
| Health consciousness | — | — | Positive | — | — | — | — | — |
| Having an acquaintance who lives alone | — | — | Not significant | — | — | — | — | — |
| Not having seen people for over a week | — | — | Positive | — | — | — | — | — |

| Factors | Adedokun et al [50] | Bergmo and Wangberg [52] | Kaga et al [55] | Ngan et al [56] | Raghu et al (1) [57] | Raghu et al (2) [57] | Shariful Islam et al [61] | Stahl et al [62] |
|---|---|---|---|---|---|---|---|---|
| **Experience with information technology** | | | | | | | | |
| Having used eHealth | — | Negative | — | — | — | — | — | — |
| Internet use | — | Not significant | — | — | — | — | — | — |
| **Attitudes** | | | | | | | | |
| Willingness to use | — | — | Positive | — | — | — | — | — |

[a] WTP: willingness to pay.

[b] The factor was not examined in the study.

[c] The favorable feature examined in the study was to involve family and friends.

**Table 8.** WTP[a]-related factors for the examined eHealth in studies that reported WTP in monthly payment.

| Factors | Cocosila et al [63] | Jacobs et al [65] | Rasche et al [66] | Somers et al (1) [34] | Somers et al (2) [34] | Tran et al [67] |
|---|---|---|---|---|---|---|
| **Characteristics of the eHealth or eHealth technology** | | | | | | |
| Favorable features | —[b] | — | Positive[c] | — | — | Positive[d] |
| **Sociodemographic characteristics** | | | | | | |
| Gender (women) | Not significant | — | Not significant | Not significant | Negative | — |
| Age | Negative | — | Not significant | Negative | Negative | — |
| Education | — | — | Not significant | — | — | Positive |
| Income | — | — | — | Positive | Not significant | — |
| Health literacy | — | — | Not significant | — | — | — |
| **Health conditions** | | | | | | |
| Perceived health status | — | — | — | Positive | Positive | — |
| Chronic conditions | — | — | Not significant | Not significant | Not significant | — |
| Health risk | — | — | Not significant | — | — | — |
| Taking regular medication | — | — | — | Not significant | Not significant | Not significant |
| Dosage of medication | — | Positive | — | — | — | — |
| **Current health care services** | | | | | | |
| Level of the health system | — | — | — | — | — | Negative |
| **Psychosocial characteristics** | | | | | | |
| Perceived autonomy support | — | Positive | — | — | — | — |
| **Experience with information technology** | | | | | | |
| Having used eHealth | — | — | — | — | — | — |
| Internet use | — | — | — | Positive and negative[e] | Not significant | — |
| SMS text messaging use | Not significant | — | — | — | — | — |
| Computer use | — | — | — | Negative[f] | Not significant | — |
| Smartphone use | — | — | — | Not significant | Not significant | — |
| Times without a mobile phone | — | — | — | — | — | Positive |
| Mobile app use | — | — | — | Not significant | Not significant | — |
| Amount spent on the phone, the internet, and additional features | — | — | — | Positive | Positive | Positive |
| Amount spent on health apps | — | — | — | Positive | Positive | — |
| **Attitudes** | | | | | | |
| Attitude toward intervention | — | — | Not significant | — | — | — |
| Ready for technology innovation | — | — | Not significant | — | — | — |
| Willingness to use | — | — | — | — | — | Positive |

[a] WTP: willingness to pay.

[b] The factor was not examined in the study.

[c] Favorable features examined in the study included decisions regarding treatment, description of physical exercise to reduce the risk of falls, continuous workout programs, and making new social contacts.

[d] Favorable features examined in the study included direct counseling with physicians and booking check-ups.

[e] Individuals who had access to the internet at home but never used it showed higher WTP than those who did not have internet access at home; individuals who had access to the internet at home and used it regularly showed lower WTP than those who did not have internet access at home.

[f] Individuals who owned a computer but rarely used it showed a lower WTP than those who did not own a computer.

## Discussion

### Principal Findings

To the best of our knowledge, this study is the first systematic review and meta-analysis of WTP for eHealth and meta-regression analysis of WTP-related factors for eHealth. We summarized and analyzed the findings of relevant scientific papers and found that the WTP value reported in each study varied significantly depending on the study population, modality used to provide eHealth, and methods used to measure WTP.

WTP for eHealth was higher in North America and Europe than in Asia and Africa, which is in line with the positive association between GDP per capita and WTP found in our meta-regression analysis. These findings suggest that even after adjusting for PPP, the overall economic condition of a country is related to people’s WTP for eHealth. Furthermore, several studies have shown that individual or household income was positively associated with WTP for eHealth in their samples, suggesting that the economic condition of an individual also predicts their WTP for eHealth. A commonly cited reason for this finding is that economic conditions affect individuals’ ability to pay, which in turn affects their WTP [77]. Another reason may be that individuals with a higher income or those in more economically developed countries have better access to and are more familiar with eHealth and have a higher intention to use and pay for it [78].

The demographic characteristics related to WTP for eHealth were gender, age, and educational level. The meta-regression analysis showed that women were associated with lower WTP values, which is in line with the findings of some studies in which women were willing to pay less than men for eHealth [34,55,61]. A possible reason for this may be that men tend to be more concerned about their health because of the higher risks of life-threatening chronic diseases than women and are more willing to pay for tools to help manage their health conditions [79,80]. Another reason may be that men tend to have a more favorable attitude toward technology than women [81] and may be more likely to accept and favor eHealth. Regarding the association between age and WTP for eHealth, there were mixed results (ie, nonsignificant, significantly positive, and significantly negative associations) among the included studies. This suggests that the association may vary drastically, depending on the context of each study (eg, population, examined eHealth, clinical setting, and alternative health services). Educational level was also related to WTP for eHealth; studies with a higher percentage of college graduates reported higher WTP values than those with a lower percentage of college graduates. This could be explained by the fact that people with higher education levels had higher eHealth literacy levels [82], perceived fewer barriers to using eHealth, and were more willing to pay for eHealth.

People were more willing to pay for eHealth provided through a specific medical device (eg, dermatoscope, nonmydriatic fundus camera, or vital sign measurement system) than for eHealth provided through websites, probably because of the advantage of obtaining accurate measurements for better clinical diagnoses. The results also showed that people were more willing to pay for eHealth provided through synchronous communication (eg, telephone calls and videoconferencing) than for health-related websites that allow for little to no interaction between users and health care providers, probably because synchronous communication enables real-time communication between users and their health care providers. Asynchronous eHealth also enables communication with health care providers through store-and-forward methods, such as SMS text messaging or email. However, the mean WTP for asynchronous eHealth was much lower than that for synchronous eHealth, probably because the timeliness of communication cannot be guaranteed through asynchronous eHealth, and the amount of health information delivered through SMS text messages or emails is limited.

The methods used to measure WTP also influenced the WTP values. Our meta-regression analysis showed that posing open-ended questions to participants resulted in lower WTP values than any other contingent valuation method. The reason may be that open-ended questions yield more 0 responses [83,84]; alternatively, answering “yes” and anchoring effects can occur when the dichotomous choice or payment scale approach is used [85]. The meta-regression analysis also revealed that ex post evaluations led to lower WTP values than ex ante evaluations, probably because individuals who had not used eHealth tended to have higher expectations and value it more. Another explanation may be that some eHealth interventions were less user-friendly [52] or failed to meet user needs in practice [86].

### Implications for Practice, Policy, and Future Research

The results of this review reflect the value of eHealth from the perspective of users, who are important sources of practical implications for the development and implementation of eHealth [87,88]. Our results showed that users place a high value on an eHealth technology that offers accurate diagnoses of health problems, has interactive features, and facilitates real-time communication with health professionals [89-91]. They also favor eHealth technology that enables shared decision-making, physical exercise training, socializing, and booking health examinations [62,66,67,92,93]. In addition, users find convenient and easy-to-use eHealth to be more attractive, suggesting that usability and technology acceptance should be taken into consideration when designing and implementing technology for eHealth, which is consistent with the literature [86,94-103].

Our results revealed the gender, education, and economic differences in the WTP for eHealth. Despite that eHealth has great potential to improve the accessibility of care by delivering health care and health information remotely and at a low cost; it might be more accessible to and create more benefits for individuals or populations that have more resources to use and are more capable of using eHealth [104]. It is a challenge for researchers, eHealth developers, and public health decision-makers to ensure that eHealth helps resolve health disparities instead of exacerbating them. We recommend identifying and removing barriers to eHealth access among disadvantaged populations [105] and keeping users’ needs and eHealth literacy levels in mind when developing eHealth interventions [106].

Our review showed that the most common approach to elicit WTP for eHealth was open-ended questions, as researchers do not have to provide cues for a reasonable WTP value, and it is easy to use. However, many participants may have never been asked these types of questions in real life and may have found it difficult to answer, leading to a low response rate and more zero responses, especially "protest zeros" [83,84]. In comparison, other formats that gave participants a starting value to consider, such as single- and double-bounded dichotomous choice models, payment scales, and bidding games, may have made it easier for them to answer the questions but could have led to anchoring bias by making the participants believe that the starting value was an appropriate value, which could have biased their responses toward that value [85]. Some studies used discrete choice experiments, in which each attribute of the good or service was valued separately instead of the full package. Discrete choice experiments generally have higher internal and external validity but require more time and effort for study design and data collection than contingent valuation studies [107]. The perfect approach for WTP evaluation remains debatable, and it seems that the approaches cannot substitute each other, which underscores the need to undertake further methodological comparisons between different approaches and explore other approaches to elicit WTP.

### Limitations

This study had some limitations that should be acknowledged. First, all the studies identified in this review were stated preference studies that used hypothetical questions to measure WTP values instead of observing actual purchases or choices made by the respondents (ie, revealed preferences). This inevitably led to a hypothetical bias, with participants reporting higher WTP values than what they would pay in real life [108-111]. The dearth of revealed preference studies in this field calls for further investigation into how much people are willing to pay for eHealth in real life and a comparison of WTP values elicited through stated preference and revealed preference methods. Second, articles written in languages other than English were excluded from this review, which may have led to language and publication bias. Third, there was great heterogeneity in the meta-analysis results, which limited the generalizability of the reported mean WTP values. Meta-analysis and meta-regression results should be interpreted with caution. Finally, we conducted a univariate meta-regression analysis as the rule of thumb is that the number of studies to be used in an analysis should be at least 10 times the number of explanatory variables in the regression [37]. Hence, this review did not use multivariate regression to control for all potential confounders and covariates when examining the associations between exploratory variables and WTP for eHealth.

### Conclusions

We found that WTP for eHealth varies greatly depending on the modality used to provide eHealth, study population, and methods used to measure WTP. We found that consumers favored and valued several eHealth modalities and features, which should be considered for adoption in future eHealth interventions. User-centered, convenient, and easy-to-use eHealth interventions should be developed, keeping in mind their usability and acceptance. Our results also showed that different population segments have significantly different WTP values for eHealth, which calls for further efforts to ensure the effective implementation of eHealth among disadvantaged populations and resolve health disparities. Thus far, there has been no consensus on the optimal approach to elicit WTP values, necessitating the exploration of other methods.

---

## Conflicts of Interest

None declared.

---

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - jmir_v24i9e25959_app1.docx ]

---

Multimedia Appendix 2
Critical appraisal of the methodological quality.
[DOCX File , 40 KB - jmir_v24i9e25959_app2.docx ]

---

## References

1. Boogerd EA, Arts T, Engelen LJ, van de Belt TH. "What is eHealth": time for an update? JMIR Res Protoc 2015 Mar 12;4(1):e29 [FREE Full text] [doi: 10.2196/resprot.4065] [Medline: 25768939]
2. Or C. Pre-implementation case studies evaluating workflow and informatics challenges in private primary care clinics for electronic medical record implementation. Int J Healthc Inf Syst Inform 2015;10(4):56-64. [doi: 10.4018/ijhisi.2015100104]
3. Or C, Tong E, Tan J, Chan S. Exploring factors affecting voluntary adoption of electronic medical records among physicians and clinical assistants of small or solo private general practice clinics. J Med Syst 2018 May 29;42(7):121. [doi: 10.1007/s10916-018-0971-0] [Medline: 29845400]

4. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
5. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
6. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
7. Or CK, Holden RJ, Valdez R. Human factors engineering and user-centered design for mobile health technology: enhancing effectiveness, efficiency, and satisfaction. In: Duffy VG, Ziefle M, Rau PL, Tseng MM, editors. Human-Automation Interaction: Mobile Computing. New York, NY, USA: Springer International Publishing; 2022.
8. Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, et al. An evaluation of the effectiveness of the modalities used to deliver electronic health interventions for chronic pain: systematic review with network meta-analysis. J Med Internet Res 2019 Jul 17;21(7):e11086 [FREE Full text] [doi: 10.2196/11086] [Medline: 31317869]
9. Slattery BW, Haugh S, Francis K, O'Connor L, Barrett K, Dwyer CP, et al. Protocol for a systematic review with network meta-analysis of the modalities used to deliver eHealth interventions for chronic pain. Syst Rev 2017 Mar 03;6(1):45 [FREE Full text] [doi: 10.1186/s13643-017-0414-x] [Medline: 28253909]
10. Chen J, Xie Z, Or C. Effectiveness of immersive virtual reality-supported interventions for patients with disorders or impairments: a systematic review and meta-analysis. Health Technol 2021 Jul 16;11(4):811-833. [doi: 10.1007/s12553-021-00561-7]
11. Verhoeven F, van Gemert-Pijnen L, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007 Dec 14;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]
12. O'Cathail M, Sivanandan MA, Diver C, Patel P, Christian J. The use of patient-facing teleconsultations in the national health service: scoping review. JMIR Med Inform 2020 Mar 16;8(3):e15380 [FREE Full text] [doi: 10.2196/15380] [Medline: 32175911]
13. de Farias FA, Dagostini CM, de Assunção Bicca Y, Falavigna VF, Falavigna A. Remote patient monitoring: a systematic review. Telemed J E Health 2020 May;26(5):576-583. [doi: 10.1089/tmj.2019.0066] [Medline: 31314689]
14. Kelly M, Fullen B, Martin D, McMahon S, McVeigh JG. eHealth interventions to support self-management in people with musculoskeletal disorders, "eHealth: it's TIME"-a scoping review. Phys Ther 2022 Apr 01;102(4):pzab307 [FREE Full text] [doi: 10.1093/ptj/pzab307] [Medline: 35079826]
15. Liu K, Xie Z, Or CK. Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020 Aug 04;8(8):e15779 [FREE Full text] [doi: 10.2196/15779] [Medline: 32459654]
16. Or CK, Tao D. Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. Int J Med Inform 2014 May;83(5):320-329. [doi: 10.1016/j.ijmedinf.2014.01.009] [Medline: 24534118]
17. Chen J, Or CK, Chen T. Effectiveness of using virtual reality-supported exercise therapy for upper extremity motor rehabilitation in patients with stroke: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jun 20;24(6):e24111 [FREE Full text] [doi: 10.2196/24111] [Medline: 35723907]
18. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
19. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
20. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5:290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
21. Or C, Tao D. A 3-month randomized controlled pilot trial of a patient-centered, computer-based self-monitoring system for the care of type 2 diabetes mellitus and hypertension. J Med Syst 2016 Apr;40(4):81. [doi: 10.1007/s10916-016-0437-1] [Medline: 26802011]
22. Or CK, Liu K, So MK, Cheung B, Yam LY, Tiwari A, et al. Improving self-care in patients with coexisting type 2 diabetes and hypertension by technological surrogate nursing: randomized controlled trial. J Med Internet Res 2020 Mar 27;22(3):e16769 [FREE Full text] [doi: 10.2196/16769] [Medline: 32217498]
23. Sülz S, van Elten HJ, Askari M, Weggelaar-Jansen AM, Huijsman R. eHealth applications to support independent living of older persons: scoping review of costs and benefits identified in economic evaluations. J Med Internet Res 2021 Mar 09;23(3):e24363 [FREE Full text] [doi: 10.2196/24363] [Medline: 33687335]
24. Bergmo TS. How to measure costs and benefits of eHealth interventions: an overview of methods and frameworks. J Med Internet Res 2015 Nov 09;17(11):e254 [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]

25. Fanta G, Pretorius L, Erasmus L. Economic analysis of sustainable eHealth implementation in developing countries: a systematic review. In: Proceedings of the 27th Annual Conference of the International Conference on Management of Technology. 2018 Presented at: IAMOT '18; April 22-26, 2018; Birmingham, UK.
26. Olsen JA, Smith RD. Theory versus practice: a review of 'willingness-to-pay' in health and health care. Health Econ 2001 Jan;10(1):39-52. [doi: 10.1002/1099-1050(200101)10:1<39::aid-hec563>3.0.co;2-e] [Medline: 11180568]
27. Robinson R. Cost-benefit analysis. BMJ 1993 Oct 09;307(6909):924-926 [FREE Full text] [doi: 10.1136/bmj.307.6909.924] [Medline: 8241859]
28. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics 2019 Feb;37(2):201-226 [FREE Full text] [doi: 10.1007/s40273-018-0734-2] [Medline: 30392040]
29. Nosratnejad S, Rashidian A, Dror DM. Systematic review of willingness to pay for health insurance in low and middle income countries. PLoS One 2016 Jun 30;11(6):e0157470 [FREE Full text] [doi: 10.1371/journal.pone.0157470] [Medline: 27362356]
30. Sunstein CR. Willingness to pay versus welfare. Harv Law Rev 2007;1:303.
31. Borghi J. Aggregation rules for cost-benefit analysis: a health economics perspective. Health Econ 2008 Jul;17(7):863-875. [doi: 10.1002/hec.1304] [Medline: 17992709]
32. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012 Feb;21(2):145-172. [doi: 10.1002/hec.1697] [Medline: 22223558]
33. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, et al. Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess 2001;5(5):1-186 [FREE Full text] [doi: 10.3310/hta5050] [Medline: 11262422]
34. Somers C, Grieve E, Lennon M, Bouamrane MM, Mair FS, McIntosh E. Valuing mobile health: an open-ended contingent valuation survey of a national digital health program. JMIR Mhealth Uhealth 2019 Jan 17;7(1):e3 [FREE Full text] [doi: 10.2196/mhealth.9990] [Medline: 30664488]
35. Buchanan J, Roope LS, Morrell L, Pouwels KB, Robotham JV, Abel L, et al. Preferences for medical consultations from online providers: evidence from a discrete choice experiment in the United Kingdom. Appl Health Econ Health Policy 2021 Jul;19(4):521-535 [FREE Full text] [doi: 10.1007/s40258-021-00642-8] [Medline: 33682065]
36. Snoswell CL, Smith AC, Page M, Caffery LJ. Patient preferences for specialist outpatient video consultations: a discrete choice experiment. J Telemed Telecare (forthcoming) 2021 Jun 18:1357633X211022898. [doi: 10.1177/1357633X211022898] [Medline: 34142895]
37. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1. London, UK: The Cochrane Collaboration; 2020.
38. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
39. Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol 2012 Sep;65(9):934-939. [doi: 10.1016/j.jclinepi.2011.11.014] [Medline: 22742910]
40. Smith RD. Construction of the contingent valuation market in health care: a critical assessment. Health Econ 2003 Aug;12(8):609-628. [doi: 10.1002/hec.755] [Medline: 12898660]
41. World Economic Outlook Database: 2021. International Monetary Fund. 2021. URL: https://www.imf.org/en/Publications/WEO/weo-database/2021/October [accessed 2022-09-05]
42. Van Houtven G. Methods for the meta-analysis of willingness-to-pay data: an overview. Pharmacoeconomics 2008;26(11):901-910. [doi: 10.2165/00019053-200826110-00003] [Medline: 18850760]
43. Higgins JP, White IR, Anzures-Cabrera J. Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. Stat Med 2008 Dec 20;27(29):6072-6092 [FREE Full text] [doi: 10.1002/sim.3427] [Medline: 18800342]
44. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
45. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
46. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
47. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
48. Zhang Y, Alonso-Coello P, Guyatt GH, Yepes-Nuñez JJ, Akl EA, Hazlewood G, et al. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences-Risk of bias and indirectness. J Clin Epidemiol 2019 Jul;111:94-104. [doi: 10.1016/j.jclinepi.2018.01.013] [Medline: 29452223]

49. Zhang Y, Coello PA, Guyatt GH, Yepes-Nuñez JJ, Akl EA, Hazlewood G, et al. GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes or values and preferences-inconsistency, imprecision, and other domains. J Clin Epidemiol 2019 Jul;111:83-93. [doi: 10.1016/j.jclinepi.2018.05.011] [Medline: 29800687]
50. Adedokun A, Idris O, Odujoko T. Patients' willingness to utilize a SMS-based appointment scheduling system at a family practice unit in a developing country. Prim Health Care Res Dev 2016 Mar;17(2):149-156. [doi: 10.1017/S1463423615000213] [Medline: 25851031]
51. Belkora J, Stupar L, O'Donnell S, Loucks A, Moore D, Jupiter C, et al. Decision support by telephone: randomized controlled trial in a rural community setting. Patient Educ Couns 2012 Oct;89(1):134-142. [doi: 10.1016/j.pec.2012.06.009] [Medline: 22776761]
52. Bergmo TS, Wangberg SC. Patients' willingness to pay for electronic communication with their general practitioner. Eur J Health Econ 2007 Jun;8(2):105-110. [doi: 10.1007/s10198-006-0014-5] [Medline: 17186205]
53. Brandling-Bennett HA, Kedar I, Pallin DJ, Jacques G, Gumley GJ, Kvedar JC. Delivering health care in rural Cambodia via store-and-forward telemedicine: a pilot study. Telemed J E Health 2005 Feb;11(1):56-62. [doi: 10.1089/tmj.2005.11.56] [Medline: 15785221]
54. Fawsitt CG, Meaney S, Greene RA, Corcoran P. Surgical site infection after caesarean section? There is an app for that: results from a feasibility study on costs and benefits. Ir Med J 2017 Sep 18;110(9):635. [Medline: 29372950]
55. Kaga S, Suzuki T, Ogasawara K. Willingness to pay for elderly telecare service using the Internet and digital terrestrial broadcasting. Interact J Med Res 2017 Oct 24;6(2):e21 [FREE Full text] [doi: 10.2196/ijmr.7461] [Medline: 29066428]
56. Ngan TT, Do VV, Huang J, Redmon PB, Minh HV. Willingness to use and pay for smoking cessation service via text-messaging among Vietnamese adult smokers, 2017. J Subst Abuse Treat 2019 Sep;104:1-6. [doi: 10.1016/j.jsat.2019.05.014] [Medline: 31370973]
57. Raghu TS, Yiannias J, Sharma N, Markus AL. Willingness to pay for teledermoscopy services at a university health center. J Patient Exp 2018 Sep;5(3):212-218 [FREE Full text] [doi: 10.1177/2374373517748657] [Medline: 30214928]
58. Ramchandran RS, Yilmaz S, Greaux E, Dozier A. Patient perceived value of teleophthalmology in an urban, low income US population with diabetes. PLoS One 2020 Jan 9;15(1):e0225300 [FREE Full text] [doi: 10.1371/journal.pone.0225300] [Medline: 31917793]
59. Rochat L, Genton B. Telemedicine for health issues while abroad: interest and willingness to pay among travellers prior to departure. J Travel Med 2018 Jan 01;25(1):tay028. [doi: 10.1093/jtm/tay028] [Medline: 29718404]
60. Ruby A, Marko-Holguin M, Fogel J, Van Voorhees BW. Economic analysis of an Internet-based depression prevention intervention. J Ment Health Policy Econ 2013 Sep;16(3):121-130 [FREE Full text] [Medline: 24327482]
61. Shariful Islam SM, Lechner A, Ferrari U, Seissler J, Holle R, Niessen LW. Mobile phone use and willingness to pay for SMS for diabetes in Bangladesh. J Public Health (Oxf) 2016 Mar;38(1):163-169. [doi: 10.1093/pubmed/fdv009] [Medline: 25687131]
62. Stahl JE, Dixon RF. Acceptability and willingness to pay for primary care videoconferencing: a randomized controlled trial. J Telemed Telecare 2010;16(3):147-151. [doi: 10.1258/jtt.2009.090502] [Medline: 20386035]
63. Cocosila M, Archer N, Yuan Y. Would people pay for text messaging health reminders? Telemed J E Health 2008 Dec;14(10):1091-1095. [doi: 10.1089/tmj.2008.0047] [Medline: 19119832]
64. Contreras-Somoza LM, Irazoki E, Castilla D, Botella C, Toribio-Guzmán JM, Parra-Vidales E, et al. Study on the acceptability of an ICT platform for older adults with mild cognitive impairment. J Med Syst 2020 May 25;44(7):120. [doi: 10.1007/s10916-020-01566-x] [Medline: 32451740]
65. Jacobs N, Drost R, Ament A, Evers S, Claes N. Willingness to pay for a cardiovascular prevention program in highly educated adults: a randomized controlled trial. Int J Technol Assess Health Care 2011 Oct;27(4):283-289. [doi: 10.1017/S0266462311000341] [Medline: 22004768]
66. Rasche P, Mertens A, Brandl C, Liu S, Buecking B, Bliemel C, et al. Satisfying product features of a fall prevention smartphone app and potential users' willingness to pay: Web-based survey among older adults. JMIR Mhealth Uhealth 2018 Mar 27;6(3):e75 [FREE Full text] [doi: 10.2196/mhealth.9467] [Medline: 29588268]
67. Tran BX, Houston S. Mobile phone-based antiretroviral adherence support in Vietnam: feasibility, patient's preference, and willingness-to-pay. AIDS Behav 2012 Oct;16(7):1988-1992. [doi: 10.1007/s10461-012-0271-5] [Medline: 22814571]
68. Tsuji M, Suzuki W, Taoka F. An empirical analysis of a telehealth system in terms of cost-sharing. J Telemed Telecare 2003;9 Suppl 1:S41-S43. [doi: 10.1258/135763303322196303] [Medline: 12952719]
69. Tsuji M, Iizuka C, Taoka F, Teshima M. The willingness of Japanese citizens to pay for e-Health systems. J Inf Technol Healthc 2006 Apr;4(2):103-110 [FREE Full text]
70. Park H, Chon Y, Lee J, Choi IJ, Yoon KH. Service design attributes affecting diabetic patient preferences of telemedicine in South Korea. Telemed J E Health 2011;17(6):442-451 [FREE Full text] [doi: 10.1089/tmj.2010.0201] [Medline: 21631382]
71. Snoswell CL, Whitty JA, Caffery LJ, Loescher LJ, Gillespie N, Janda M. Direct-to-consumer mobile teledermoscopy for skin cancer screening: preliminary results demonstrating willingness-to-pay in Australia. J Telemed Telecare 2018 Dec;24(10):683-689. [doi: 10.1177/1357633X18799582] [Medline: 30343653]
72. Spinks J, Janda M, Soyer HP, Whitty JA. Consumer preferences for teledermoscopy screening to detect melanoma early. J Telemed Telecare 2016 Jan;22(1):39-46. [doi: 10.1177/1357633X15586701] [Medline: 26026184]

73. van der Pol M, McKenzie L. Costs and benefits of tele-endoscopy clinics in a remote location. J Telemed Telecare 2010;16(2):89-94. [doi: 10.1258/jtt.2009.090609] [Medline: 20139140]
74. Ahn J, Shin J, Lee J, Shin K, Park H. Consumer preferences for telemedicine devices and services in South Korea. Telemed J E Health 2014 Feb;20(2):168-174. [doi: 10.1089/tmj.2013.0101] [Medline: 24303932]
75. Chang J, Savage SJ, Waldman DM. Estimating willingness to pay for online health services with discrete-choice experiments. Appl Health Econ Health Policy 2017 Aug;15(4):491-500. [doi: 10.1007/s40258-017-0316-z] [Medline: 28290107]
76. Deal K, Keshavjee K, Troyan S, Kyba R, Holbrook AM. Physician and patient willingness to pay for electronic cardiovascular disease management. Int J Med Inform 2014 Jul;83(7):517-528. [doi: 10.1016/j.ijmedinf.2014.04.007] [Medline: 24862891]
77. Donaldson C. Valuing the benefits of publicly-provided health care: does 'ability to pay' preclude the use of 'willingness to pay'? Soc Sci Med 1999 Aug;49(4):551-563. [doi: 10.1016/s0277-9536(99)00173-2] [Medline: 10414814]
78. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
79. Rieker PP, Bird CE. Rethinking gender differences in health: why we need to integrate social and biological perspectives. J Gerontol B Psychol Sci Soc Sci 2005 Oct;60 Spec No 2:40-47. [doi: 10.1093/geronb/60.special_issue_2.s40] [Medline: 16251589]
80. Chung WS, Shin KO, Bae JY. Gender differences in body image misperception according to body mass index, physical activity, and health concern among Korean university students. J Mens Health 2019 Jan 22;15(1):e1-e9. [doi: 10.22374/1875-6859.15.1.1]
81. Cai Z, Fan X, Du J. Gender and attitudes toward technology use: a meta-analysis. Comput Educ 2017 Feb;105:1-13. [doi: 10.1016/j.compedu.2016.11.003]
82. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
83. Donaldson C, Thomas R, Torgerson DJ. Validity of open-ended and payment scale approaches to eliciting willingness to pay. Appl Econ 1997;29(1):79-84. [doi: 10.1080/000368497327425]
84. Whynes DK, Frew E, Wolstenholme JL. A comparison of two methods for eliciting contingent valuations of colorectal cancer screening. J Health Econ 2003 Jul;22(4):555-574. [doi: 10.1016/S0167-6296(03)00006-7] [Medline: 12842315]
85. Frew EJ, Whynes DK, Wolstenholme JL. Eliciting willingness to pay: comparing closed-ended with open-ended and payment scale formats. Med Decis Making 2003;23(2):150-159. [doi: 10.1177/0272989X03251245] [Medline: 12693877]
86. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
87. Hensher M, Cooper P, Dona SW, Angeles MR, Nguyen D, Heynsbergh N, et al. Scoping review: development and assessment of evaluation frameworks of mobile health apps for recommendations to consumers. J Am Med Inform Assoc 2021 Jun 12;28(6):1318-1329 [FREE Full text] [doi: 10.1093/jamia/ocab041] [Medline: 33787894]
88. Martin-Payo R, Carrasco-Santos S, Cuesta M, Stoyan S, Gonzalez-Mendez X, Fernandez-Alvarez MD. Spanish adaptation and validation of the User Version of the Mobile Application Rating Scale (uMARS). J Am Med Inform Assoc 2021 Nov 25;28(12):2681-2686. [doi: 10.1093/jamia/ocab216] [Medline: 34613400]
89. Runz-Jørgensen SM, Schiøtz ML, Christensen U. Perceived value of eHealth among people living with multimorbidity: a qualitative study. J Comorb 2017 Aug 24;7(1):96-111 [FREE Full text] [doi: 10.15256/joc.2017.7.98] [Medline: 29359124]
90. Hersh WR, Hickam DH, Severance SM, Dana TL, Pyle Krages K, Helfand M. Diagnosis, access and outcomes: update of a systematic review of telemedicine services. J Telemed Telecare 2006;12 Suppl 2:S3-31. [doi: 10.1258/135763306778393117] [Medline: 16989671]
91. Akesson KM, Saveman BI, Nilsson G. Health care consumers' experiences of information communication technology--a summary of literature. Int J Med Inform 2007 Sep;76(9):633-645. [doi: 10.1016/j.ijmedinf.2006.07.001] [Medline: 16931133]
92. Chen T, Or CK. Development and pilot test of a machine learning-based knee exercise system with video demonstration, real-time feedback, and exercise performance score. Proc Hum Factors Ergon Soc Annu Meet 2021 Nov 12;65(1):1519-1523. [doi: 10.1177/1071181321651109]
93. Chan KL, Leung WC, Tiwari A, Or KL, Ip P. Using smartphone-based psychoeducation to reduce postnatal depression among first-time mothers: randomized controlled trial. JMIR Mhealth Uhealth 2019 May 14;7(5):e12794 [FREE Full text] [doi: 10.2196/12794] [Medline: 31094354]
94. Dopp AR, Parisi KE, Munson SA, Lyon AR. Aligning implementation and user-centered design strategies to enhance the impact of health services: results from a concept mapping study. Implement Sci Commun 2020 Feb 26;1:17 [FREE Full text] [doi: 10.1186/s43058-020-00020-w] [Medline: 32885179]
95. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 8;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
96. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 08;1(2):e13 [FREE Full text] [doi: 10.2196/resprot.2184] [Medline: 23612015]

97. Karsh BT, Holden RJ, Or CK. Human factors and ergonomics of health information technology implementation. In: Carayon P, editor. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. 2nd edition. Boca Raton, FL, USA: CRC Press; 2011:249-264.
98. Liu K, Or CK, So M, Cheung B, Chan B, Tiwari A, et al. A longitudinal examination of tablet self-management technology acceptance by patients with chronic diseases: integrating perceived hand function, perceived visual function, and perceived home space adequacy with the TAM and TPB. Appl Ergon 2022 Apr;100:103667. [doi: 10.1016/j.apergo.2021.103667] [Medline: 34920356]
99. Or CK, Valdez RS, Casper GR, Carayon P, Burke LJ, Brennan PF, et al. Human factors and ergonomics in home care: current concerns and future considerations for health information technology. Work 2009;33(2):201-209 [FREE Full text] [doi: 10.3233/WOR-2009-0867] [Medline: 19713630]
100. Xie Z, Kalun Or C. Acceptance of mHealth by elderly adults: a path analysis. Proc Hum Factors Ergon Soc Annu Meet 2021 Feb 09;64(1):755-759. [doi: 10.1177/1071181320641174]
101. Yan M, Or C. Factors in the 4-week acceptance of a computer-based, chronic disease self-monitoring system in patients with type 2 diabetes mellitus and/or hypertension. Telemed J E Health 2018 Feb;24(2):121-129. [doi: 10.1089/tmj.2017.0064] [Medline: 28737995]
102. K.L. C, Karsh BT. The patient technology acceptance model (PTAM) for homecare patients with chronic illness. Proc Hum Factors Ergon Soc Annu Meet 2006 Oct 1;50(10):989-993. [doi: 10.1177/154193120605001040]
103. Scheibner J, Sleigh J, Ienca M, Vayena E. Benefits, challenges, and contributors to success for national eHealth systems implementation: a scoping review. J Am Med Inform Assoc 2021 Aug 13;28(9):2039-2049 [FREE Full text] [doi: 10.1093/jamia/ocab096] [Medline: 34151990]
104. Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. Am J Prev Med 2007 May;32(5 Suppl):S131-S133 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.012] [Medline: 17466818]
105. Mangin D, Parascandalo J, Khudoyarova O, Agarwal G, Bismah V, Orr S. Multimorbidity, eHealth and implications for equity: a cross-sectional survey of patient perspectives on eHealth. BMJ Open 2019 Feb 12;9(2):e023731 [FREE Full text] [doi: 10.1136/bmjopen-2018-023731] [Medline: 30760515]
106. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
107. Breidert C, Hahsler M, Reutterer T. A review of methods for measuring willingness-to-pay. Innov Mark 2006;2(4):8-32. [doi: 10.1007/978-3-8350-9244-0_4]
108. Voelckner F. An empirical comparison of methods for measuring consumers’ willingness to pay. Market Lett 2006 Apr;17(2):137-149. [doi: 10.1007/s11002-006-5147-x]
109. Botelho A, Pinto LC. Hypothetical, real, and predicted real willingness to pay in open-ended surveys: experimental results. Appl Econ Lett 2002 Dec;9(15):993-996. [doi: 10.1080/13504850210148143]
110. Wertenbroch K, Skiera B. Measuring consumers' willingness to pay at the point of purchase. J Mark Res 2002 May 1;39(2):228-241. [doi: 10.1509/jmkr.39.2.228.19086]
111. Kanya L, Sanghera S, Lewin A, Fox-Rushby J. The criterion validity of willingness to pay methods: a systematic review and meta-analysis of the evidence. Soc Sci Med 2019 Jul;232:238-261 [FREE Full text] [doi: 10.1016/j.socscimed.2019.04.015] [Medline: 31108330]

## Abbreviations

**GDP:** gross domestic product
**GRADE:** Grading of Recommendations, Assessment, Development, and Evaluation
**PPP:** purchasing power parity
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**WTP:** willingness to pay

*Edited by A Mavragani; submitted 22.11.20; peer-reviewed by A Iftikhar, M Behzadifar, MS Kim; comments to author 07.01.21; revised version received 15.06.22; accepted 11.08.22; published 14.09.22.*

*Please cite as:*
*Xie Z, Chen J, Or CK*
*Consumers’ Willingness to Pay for eHealth and Its Influencing Factors: Systematic Review and Meta-analysis*
*J Med Internet Res 2022;24(9):e25959*
*URL: https://www.jmir.org/2022/9/e25959*
*doi:10.2196/25959*
*PMID:36103227*

Review

# Comparability of Patients in Trials of eHealth and Face-to-Face Psychotherapeutic Interventions for Depression: Meta-synthesis

Vera Aemissegger[1], MD; Jesus Lopez-Alcalde[1,2,3], MD; Claudia M Witt[1,4,5], Prof Dr; Jürgen Barth[1], PhD

[1]Institute for Complementary and Integrative Medicine, University Hospital Zurich and University of Zurich, Zurich, Switzerland

[2]Faculty of Health Sciences, Universidad Francisco de Vitoria, Madrid, Spain

[3]Unidad de Bioestadística Clínica, Hospital Universitario Ramón y Cajal, IRYCIS, CIBERESP, Madrid, Spain

[4]Berlin Institute of Health, Institute of Social Medicine, Epidemiology and Health Economics, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany

[5]Center for Integrative Medicine, University of Maryland School of Medicine, Baltimore, MD, United States

**Corresponding Author:**
Jürgen Barth, PhD
Institute for Complementary and Integrative Medicine
University Hospital Zurich and University of Zurich
Sonneggstrasse 6
Zurich, 8091
Switzerland
Phone: 41 44 255 48 96
Email: mail@juergen-barth.de

## Abstract

**Background:** Depressive disorders (DDs) are a public health problem. Face-to-face psychotherapeutic interventions are a first-line option for their treatment in adults. There is a growing interest in eHealth interventions to maximize accessibility for effective treatments. Thus, the number of randomized controlled trials (RCTs) of eHealth psychotherapeutic interventions has increased, and these interventions are being offered to patients. However, it is unknown whether patients with DDs differ in internet-based and face-to-face intervention trials. This information is essential to gain knowledge about eHealth trials' external validity.

**Objective:** We aimed to compare the baseline characteristics of patients with DDs included in the RCTs of eHealth and face-to-face psychotherapeutic interventions with a cognitive component.

**Methods:** In this meta-epidemiological study, we searched 5 databases between 1990 and November 2017 (MEDLINE, Embase, PsycINFO, Google Scholar, and the database of Cuijpers et al). We included RCTs of psychotherapeutic interventions with a cognitive component (eg, cognitive therapy, cognitive behavioral therapy [CBT], or interpersonal therapy) delivered face-to-face or via the internet to adults with DDs. Each included study had a matching study for predefined criteria to allow a valid comparison of characteristics and was classified as a face-to-face (CBT) or eHealth (internet CBT) intervention trial. Two authors selected the studies, extracted data, and resolved disagreements by discussion. We tested whether predefined baseline characteristics differed in face-to-face and internet-based trials using a mixed-effects model and testing for differences with *z* tests (statistical significance set at .05). For continuous outcomes, we also estimated the difference in means between subgroups with 95% CI.

**Results:** We included 58 RCTs (29 matching pairs) with 3846 participants (female: n=2803, 72.9%) and mean ages ranging from 20-74 years. White participants were the most frequent (from 63.6% to 100%). Other socioeconomic characteristics were poorly described. The participants presented DDs of different severity measured with heterogeneous instruments. Internet CBT trials had a longer depression duration at baseline (7.19 years higher, CI 95% 2.53-11.84; 10.0 vs 2.8 years; *P*=.002), but the proportion of patients with previous depression treatment was lower (24.8% vs 42%; *P*=.04). Subgroup analyses found no evidence of differences for the remaining baseline characteristics: age, gender, education, living area, depression severity, history of depression, actual antidepressant medication, actual physical comorbidity, actual mental comorbidity, study dropout, quality of life, having children, family status, and employment. We could not compare proficiency with computers due to the insufficient number of studies.

**Conclusions:** The baseline characteristics of patients with DDs included in the RCTs of eHealth and face-to-face psychotherapeutic interventions are generally similar. However, patients in eHealth trials had a longer duration of depression, and a lower proportion had received previous depression treatment, which might indicate that eHealth trials attract patients who postpone earlier treatment attempts.

**Trial Registration:** PROSPERO CRD42019085880; https://tinyurl.com/4xufwcyr

***(J Med Internet Res 2022;24(9):e36978)*** doi:10.2196/36978

**KEYWORDS**

depression; mental health; digital intervention; eHealth; web-based; randomized controlled trial; RCT; meta-analysis; epidemiology; epidemiological; depressive disorder; mental illness; mental condition; mental disorder; psychotherapy; psychotherapeutic intervention; CBT; iCBT; cognitive behavioral therapy; face-to-face; cognitive therapy; interpersonal therapy

## Introduction

Depressive disorders (DDs) affect more than 300 million people worldwide and have prevailed as a leading nonfatal health issue for almost 3 decades [1,2]. Their incidence and burden are expected to rise in the future. At present, DDs cause 15% of the days lived with disability and is associated with suicide; a high burden in personal, family, social, and work life; and elevated health care costs [2,3]. Increasing the accessibility to effective interventions for DDs is an international priority [4,5].

Psychotherapeutic interventions aim to improve depressive symptoms by increasing self-efficacy, developing coping skills, and changing cognitions, emotions, and behaviors with exercises and sometimes homework between sessions. Examples include cognitive therapy, cognitive behavioral therapy (CBT), interpersonal therapy, and psychodynamic treatments. Face-to-face psychotherapeutic interventions are accepted as a first-line treatment for DDs [6-9], and the different approaches have benefits from small to large magnitude compared to usual care (standardized mean differences [SMDs] ranging from –0.32 to –0.92) [10,11].

Internet-based (also known as eHealth) psychotherapeutic interventions, such as internet CBT (iCBT), treat psychological problems via digital platforms [12]. The eHealth approach involves adapting standard face-to-face protocols into computerized self-help material delivered over a period either as a self-help program alone or combined with brief therapist support [4]. Similar to face-to-face interventions, most randomized controlled trials (RCTs) in eHealth psychotherapy have evaluated interventions with CBT components [6].

Available meta-analyses suggest that internet-based psychotherapeutic interventions are effective for DDs compared to a waiting list or attention control condition. Internet-based interventions improved depression severity in adults with major depression (SMD –0.90, 95% CI –1.07 to –0.73) [13]. Self-guided iCBT was effective on depressive symptom severity (SMD 0.27) and treatment response (odds ratio 1.95, 95% CI 1.52-2.50) [14]. Internet-based interventions could also be effective for treating subthreshold depression and preventing major depression [15]. Recent meta-analyses of direct comparisons suggest similar effectiveness between internet-based and face-to-face psychotherapeutic interventions for treating DDs [6,13,16-18], which was the starting point of this project.

Although the previous data suggest that iCBT can be as effective as face-to-face CBT for treating DDs, the evidence is not conclusive. Other systematic reviews found that iCBT is more effective than face-to-face CBT at reducing symptom severity in depression (SMD –1.73, 95% CI –2.72 to –0.74) [19], but there was no evidence of an effect of iCBT compared to active treatments regarding depressive symptoms in adolescents and young adults [20]. Conversely, heterogeneity between studies was substantial in the published meta-analyses, with *I*-squared ($I^2$) up to 98%. In addition, no moderators of treatment effects were identified [20] that explained this heterogeneity. Thus, the comparative effectiveness of eHealth and face-to-face interventions is unclear.

At present, using technology to maximize accessibility for depression treatments is an important next step [6]. Digital interventions can increase access to evidence-based psychotherapeutic interventions, which explains the large-scale investments to integrate eHealth into regular health care services [4,6,21]. This context requires understanding under what conditions eHealth interventions work, if they are effective for everyone, and in which groups the interventions might be more or less effective [22,23]. For example, there are barriers to using iCBT for patients, such as computer literacy, language, and disabilities, and patients' attrition rates can differ between face-to-face and digital interventions [24]. Additionally, psychologists must be trained to deliver eHealth interventions [6]. Finally, recommendations for using iCBT should be underpinned by high-quality evidence from studies including complex presentations of depression—that is, patients with comorbidities, higher depression severity, or risk of suicide. In summary, the evidence on digital psychotherapeutic interventions should apply to everyday practice and all patient groups [6].

The baseline characteristics of patients in the RCTs of eHealth interventions for DDs have received little consideration. Determining whether these characteristics differ among eHealth and face-to-face intervention trials is essential to gain knowledge about the external validity of eHealth trials. The aim of our study was to compare the baseline characteristics of patients with DDs included in the RCTs of eHealth and face-to-face psychotherapeutic interventions with a cognitive component.

## Methods

### Registration

This meta-epidemiological study was prospectively registered in PROSPERO (registration CRD42019085880).

## Inclusion Criteria

### *Date*

We included RCTs published as an article in any language from 1990 to November 2017.

### *Participants*

Participants included adults (aged ≥16 years) with a diagnosis of DD according to an established diagnostic procedure. Depression could be the only diagnosis or coexist with other conditions, but DD should be the leading psychological diagnosis. We excluded studies with patients who are hospitalized.

### *Interventions*

Psychotherapeutic interventions with at least 5 sessions and a cognitive component—that is, cognitive therapy, CBT, and interpersonal therapy—were eligible. We tried to reduce the heterogeneity among the included psychotherapeutic interventions by focusing on those with a cognitive component. From now on, we will label these interventions as CBT, since they share basic principles, such as that cognitions contribute to the maintenance of depression via their association with emotions and behaviors [25]. We excluded supportive therapy, psychodynamic treatment, and interventions delivered at group level (group, family, or couple therapies).

As a comparator, the studies should have another psychotherapeutic intervention, a sham intervention, or an inactive control (such as a waiting list). To reduce heterogeneity among the included studies, we excluded pharmacological treatment or bibliotherapy as comparators, since the motivation to participate in these trials may differ. However, we admitted antidepressants with stable dosage as cointervention, as the combination of antidepressants and psychotherapeutic interventions reflects routine practice in managing DDs. We created 2 subgroups of studies based on the following criteria.

1. eHealth CBT interventions (iCBT): This group included studies evaluating the effects of an eHealth CBT intervention (internet- or device-based self-help program delivered via computer or smartphone). The iCBT must be provided by a health professional with minimum or absent email support. We excluded studies with regular or direct web-based contact (eg, web-based session or chat) or using bibliotherapy on screen. We acknowledge that these can also be eHealth interventions, but we focused on interventions requiring patients working on their own.
2. Face-to-face CBT interventions (CBT): This group included studies evaluating the effects of a CBT intervention delivered face-to-face—that is, the sessions require the patient and therapist being in the same room with direct contact. We excluded interventions delivered without visual contact—for example, communication via chat or phone exclusively.

### *Outcomes*

We compared the patients' characteristics at baseline, as shown in Textbox 1.

**Textbox 1.** Primary and secondary outcomes.

**Primary outcomes**

- Age (years; mean, SD)
- Gender (proportion of women)
- Education (proportion of patients with higher education; ie, at least a high school degree)
- Living area (proportion of patients living in a metropolitan area)
- Depression score (mean, SD)
- Depression duration (years; mean, SD)
- History of depression (proportion of patients with at least one previous episode of depression)
- Previous depression treatment (proportion of patients who had received any kind of treatment for depression; ie, psychotherapy, antidepressants, or both)
- Actual antidepressant medication (proportion of patients receiving antidepressants at the start of [and during] the study)
- Actual physical comorbidity (proportion of patients having at least one physical comorbidity; eg, diabetes mellitus)
- Actual mental comorbidity (proportion of patients having at least one additional mental disorder; eg, Axis I diagnosis)
- Study dropout (proportion of patients who dropped out or did not finish the study)

**Secondary outcomes**

- Quality of life (measured with a validated scale; mean, SD)
- Proficiency with computers (measured with a validated scale; mean, SD)
- Having children (proportion of patients having children)
- Family status (proportion of patients living alone; ie, single, divorced, or widowed)
- Employment (proportion of patients being employed)

### Search Methods for the Identification of Studies

First, we searched the following sources for face-to-face CBT intervention studies: (1) the database of Cuijpers et al [26] (date consulted: November 11, 2017), which was created in 2009, contains the RCTs of psychological treatments for depression, and on the date of our search (November 11, 2017), consisted of a total of 351 records; and (2) the collection of articles at the Institute for Complementary and Integrative Medicine (University Hospital Zurich and University of Zurich, Switzerland).

Second, we searched the following electronic databases (from October 16 to December 31, 2017) for iCBT studies: (1) MEDLINE (via PubMed), (2) Embase, (3) PsycINFO, and (4) Google Scholar. The search strategies combined relevant search terms related to the main concepts of the search (depression, eHealth, and RCTs). We also screened the bibliographies of key publications (see Multimedia Appendix 1).

### Selection of Studies

In the first stage, 1 author (VA) screened the records (titles and abstracts) in the database of Cuijpers et al [26] to select potentially eligible face-to-face CBT studies. VA also screened the results of the electronic searches and the bibliographies of key publications for iCBT studies that could possibly be matched with the face-to-face CBT studies. If there was any uncertainty based on the information given in the title or abstract, VA asked another author (JB) for clarification. Subsequently, 2 authors assessed the eligibility of the retrieved full texts: VA assessed the eligibility of each full text, which was cross-checked by JB. Disagreements were resolved by discussion. In the second stage, 2 researchers (VA and Lena Kümmel) independently checked if the potentially eligible studies reported all the necessary data. VA and Lena Kümmel resolved disagreements by discussion, and if there was no consensus, JB reached a final decision.

To include a study in the analysis, it must have had a matching study (being either iCBT or face-to face CBT) for all the following factors (all of them predefined and implemented in this order): (1) the same depression measurement or scale (eg, Edinburgh Postnatal Depression Scale), (2) similar depression treatment (eg, CBT), (3) similar diagnosis, (4) similar age range, (5) similar recruitment timeframe (less than 5 years of difference), (6) similar publication dates (less than 5 years of difference), (7) similar country, and (8) similar race. VA implemented the matching, and JB checked the decisions.

### Data Extraction

Next, 2 researchers (Lena Kümmel and VA) independently used a Microsoft Excel form to extract data on participants, interventions, comparators, outcomes, and matching criteria. JLA cross-checked the extracted numerical data. Discrepancies were resolved by discussion. We did not assess the risk of bias or contact the study authors to clarify unclear information.

For each outcome, we extracted the total number of randomized participants, the number of participants with the characteristic (dichotomous data), and the mean and SD (continuous data). If different scales were used for the same construct, we standardized each study's mean and SD to a 100-point scale. To standardize means, we applied the following formula: *(mean – lowest scale value)× 100/scale range*. To standardize the SDs, we multiplied the observed SD by 100 and divided it by the scale range. As this standardization method does not correct for differences in the direction of the scales, we checked that all the scales pointed in the same direction; for example, if higher values indicated a better health state. If different time units were used, we converted the time unit to years.

For missing SDs, we first tried to calculate them from the report using the Review Manager calculator (version 5.4.1; The Cochrane Collaboration) [27]. We followed the method of Wan et al [28] to estimate the mean and SD from the sample size, median, range, and IQR. If these procedures were not possible, we borrowed the SD from other studies in the same meta-analysis [29]. If several candidate SDs were available, we used the largest SD.

We attempted to perform an "available-case analysis" of the randomized population: we took as denominators the randomized participants with a complete baseline measurement of the outcome. We considered the randomized population if the population measured at baseline was unclear. When authors presented the baseline information for those who completed the intervention and those lost to follow-up separately, we pooled the data with the Review Manager calculator (version 5.4.1; The Cochrane Collaboration) [27].

### Statistical Analysis

We used the Comprehensive Meta-Analysis software (version 3; Biostat) [30] to perform the analyses. We applied a DerSimonian and Laird random-effects meta-analysis [31]. The logit transformation was used for meta-analyses of prevalence. We investigated statistical heterogeneity in the results by considering the chi-square *P* value and the $I^2$ statistic [32]. When the relevant levels of heterogeneity were found, we still meta-analyzed the data.

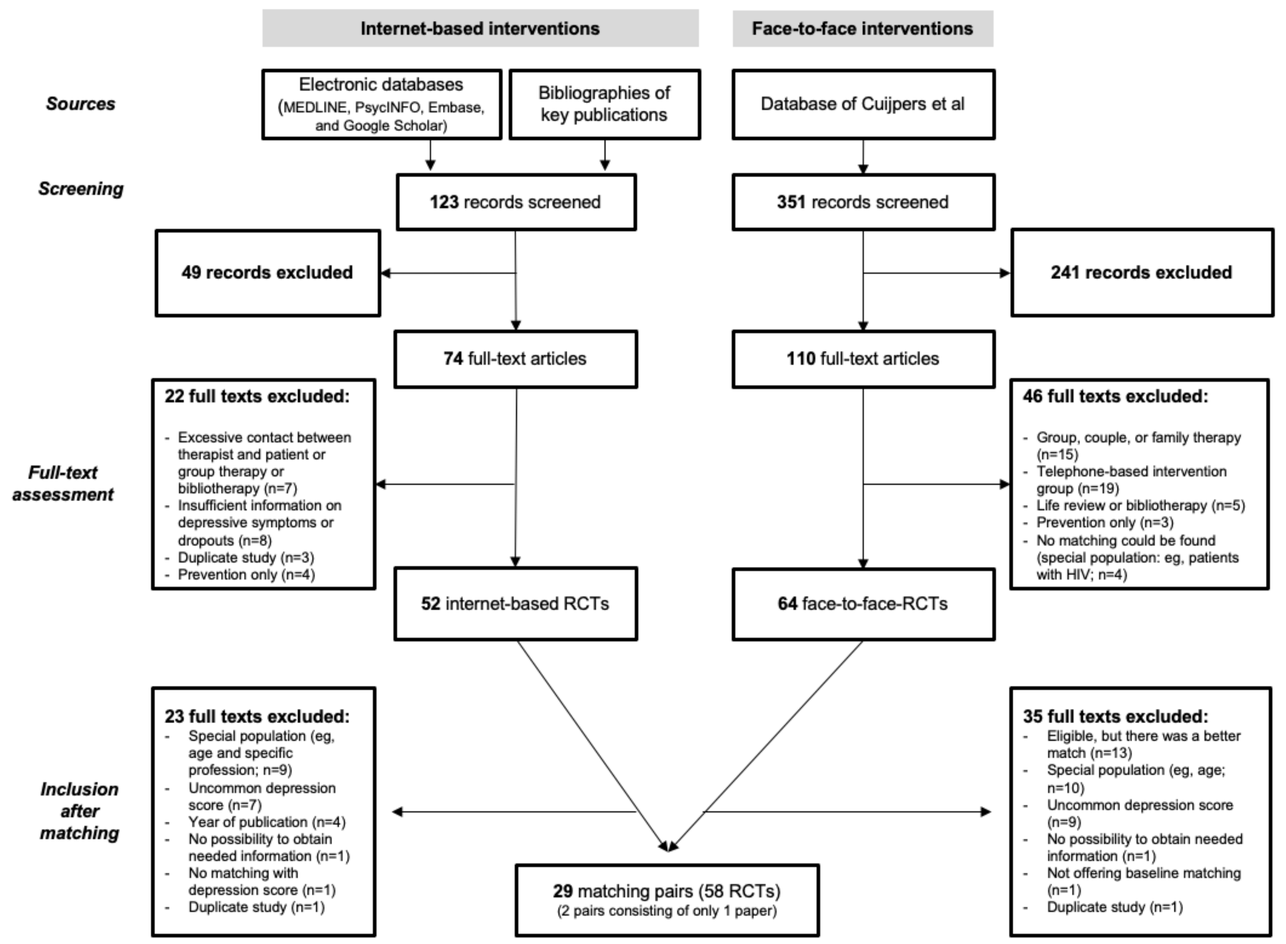
To test whether each baseline characteristic differed in CBT and iCBT trials, we used a mixed-effects model. This model pools the studies within each subgroup using the random-effects model and tests for differences between the subgroups using a fixed-effects model [33]. Due to the low number of studies (below 30 per subgroup), we used a common among-study variance ($\tau^2$) for each subgroup, which was computed by pooling the among-study variances of the 2 subgroups [33]. We ran a *z* test to compare the 2 effect sizes directly. The threshold for statistical significance was .05. For continuous outcomes, we also estimated the difference in means between subgroups along with its 95% CI [33]. We did not obtain a 95% CI for the difference in prevalence as there is no meaningful way to compute it [33].

## Results

### Search Result

The search for iCBT studies generated 123 records, and the search for face-to-face studies found 351 records. Therefore, we screened 474 titles and abstracts and excluded 290. We examined 184 full-text reports, of which 68 were excluded. We further assessed 64 face-to-face and 52 internet-based full-text RCTs, from which we finally included 29 matching pairs (with a total of 58 included RCTs). More details are provided in Figure 1.

**Figure 1.** Flow chart. RCT: randomized controlled trial. Face-to-face studies screened from the database of Cuijpers et al [26].



## Description of the Included Studies

Multimedia Appendices 2 and 3 detail the characteristics of the 58 included studies, which were published between 1990 and 2017 (n=52, 90% of them after 2000). All studies were carried out in high-income countries: 50% (n=29) in Europe, 22% (n=13) in the United States, 17% (n=10) in Australia, and 10% (n=6) in other countries. There were no multinational studies.

The studies included a total of 3846 patients. The patients were adults (mean ages ranging from 20 to 74 years in the 57 studies reporting this information) and mostly female (n=2803, 72.9%). In the 29 studies reporting the patients’ race, White patients were the most frequently reported, representing from 63.6% to 100% of the samples. The participants’ socioeconomic status was poorly described in the 58 studies: 39 (67%) studies reported the participants’ education, ranging from college to doctoral degrees. There were 27 (47%) studies that reported the participants’ employment status: from 14% to 80% of the patients were employed (full- or part-time). There were 4 (7%) studies that reported the social class or income of the included participants: from 16% to 36% of the participants had a social class I/II or an income higher than US $30,000/year.

The included patients had different types of DDs: mild to moderate, major depression, postnatal depression, and others. Depression severity was measured with different tools, with the Beck Depression Index being the most common (n=20, 34%). The participants presented mental (eg, addictions) and physical (eg, diabetes mellitus or cardiac surgery) comorbidities that were matched in the study pairs.

All the included studies delivered psychotherapeutic interventions with a cognitive component. The duration of the interventions ranged from 6 to 20 weeks with daily, weekly, or fortnightly sessions that lasted from 10 to 90 minutes each. The cointerventions were poorly described.

## Comparison of Baseline Characteristics of Participants in the Included iCBT and Face-to-Face CBT RCTs

The results are provided in Table 1. Multimedia Appendix 4 also details the results of the meta-analyses, including the iCBT and face-to-face CBT studies.

**Table 1.** Comparison of eHealth and face-to-face psychotherapeutic studies according to baseline characteristics. No study provided data for the outcome "proficiency with computers."

| Characteristic | Meta-analyses (random-effects model) | | | | | |
|---|---|---|---|---|---|---|
| | Estimate (95% CI) | Study, n | Participant, n | $I^2$ (%) | Subgroup analyses[a] | |
| | | | | | Difference (95% CI)[b] | *P* value |
| **Age (years)** | | | | | –1.89 (–10.08 to 6.29) | .65 |
| iCBT[c] | 39.98 (35.22-44.74) | 29 | 2574 | 99.7 | | |
| Face-to-face CBT[d] | 41.81 (35.21-48.54) | 28 | 1056 | 99.5 | | |
| **Gender, women, %** | | | | | —[e] | .16 |
| iCBT | 72.9 (69.2-76.4) | 29 | 2575 | 66 | | |
| Face-to-face CBT | 68.2 (62.2-73.7) | 29 | 1271 | 71.1 | | |
| **Education, higher education, %** | | | | | — | .38 |
| iCBT | 84.1 (77.8-88.8) | 22 | 1801 | 86.4 | | |
| Face-to-face CBT | 79.2 (67.4-87.5) | 15 | 789 | 85.2 | | |
| **Living area, metropolitan area, %** | | | | | — | .38 |
| iCBT | 99.5 (96.2-99.9) | 2 | 345 | 36.9 | | |
| Face-to-face CBT | 98.1 (88-99.7) | 2 | 52 | <0.001 | | |
| **Depression score[f] (standardized to a 0-100–point scale)** | | | | | 1.10 (–3.43 to 5.61) | .64 |
| iCBT | 41.34 (37.37-45.31) | 29 | 2581 | 98.6 | | |
| Face-to-face CBT | 42.25 (38.09-42.41) | 28 | 1020 | 96.5 | | |
| **Depression duration (years)** | | | | | 7.19 (2.53-11.84) | .002 |
| iCBT | 10.0 (5.6-14.4) | 1 | 36 | 0 | | |
| Face-to-face CBT | 2.8 (1.2-4.4) | 5 | 155 | 89.2 | | |
| **History of depression, %** | | | | | — | .42 |
| iCBT | 56.6 (39-72.7) | 10 | 774 | 93.3 | | |
| Face-to-face CBT | 65.1 (53.1-75.5) | 10 | 342 | 73.3 | | |
| **Previous depression treatment[g], %** | | | | | — | .04 |
| iCBT | 24.8 (18-33.1) | 8 | 908 | 75.2 | | |
| Face-to-face CBT | 42 (28.3-57.1) | 7 | 303 | 80.9 | | |
| **Actual antidepressant medication, %** | | | | | — | .11 |
| iCBT | 33.1 (23.6-44.2) | 13 | 1419 | 91.3 | | |
| Face-to-face CBT | 14.8 (5-36.6) | 13 | 423 | 85.3 | | |
| **Actual physical comorbidity, %** | | | | | — | .33 |
| iCBT | 99.6 (97.3-99.9) | 2 | 254 | 0 | | |
| Face-to-face CBT | 98.5 (90-99.8) | 2 | 66 | 0 | | |
| **Actual mental comorbidity, %** | | | | | — | .77 |
| iCBT | 73.8 (39.2-92.5) | 5 | 132 | 84.6 | | |
| Face-to-face CBT | 66.9 (28.7-91.1) | 5 | 196 | 89.7 | | |
| **Study dropout, %** | | | | | — | .36 |
| iCBT | 19.5 (14.1-26.4) | 24 | 1878 | 89.5 | | |
| Face-to-face CBT | 15.4 (10.1-22.7) | 24 | 987 | 83.4 | | |
| **Quality of life (standardized to a 0-100–point scale)** | | | | | 14.50 (–12.54 to 41.53) | .29 |
| iCBT | 48.11 (36.5-59.62) | 9 | 904 | 99.3 | | |

| Characteristic | Meta-analyses (random-effects model) | | | | | |
|---|---|---|---|---|---|---|
| | Estimate (95% CI) | Study, n | Participant, n | $I^2$ (%) | Subgroup analyses[a] | |
| | | | | | Difference (95% CI)[b] | *P* value |
| Face-to-face CBT | 33.61 (9.15-58.07) | 2 | 90 | 98.5 | | |
| **Having children, %** | | | | | — | .55 |
| iCBT | 99 (95.3-99.8) | 3 | 221 | 0 | | |
| Face-to-face CBT | 98.1 (91-99.6) | 3 | 79 | 0 | | |
| **Family status, living alone, %** | | | | | — | .37 |
| iCBT | 38.3 (30.8-46.5) | 20 | 1795 | 88.2 | | |
| Face-to-face CBT | 44.2 (34.5-54.4) | 20 | 768 | 83.3 | | |
| **Employment, %** | | | | | — | .45 |
| iCBT | 59.4 (47.9-69.9) | 13 | 1413 | 90.6 | | |
| Face-to-face CBT | 53 (40.9-64.8) | 13 | 519 | 80.9 | | |

[a] Degrees of freedom=1.

[b] 95% CI for the difference in prevalence was not calculated, as there is no meaningful way to compute it.

[c] iCBT: internet cognitive behavioral therapy.

[d] CBT: cognitive behavioral therapy.

[e] Not available.

[f] Subgroup analyses for depression measured with individual scores: Beck Depression Inventory (*P*≥.99); Depression, Anxiety and Stress Scale-21 (*P*=.04); Center for Epidemiologic Studies-Depression Scale (6 studies; *P*=.87), Edinburgh Postnatal Depression Scale (4 studies; *P*=.34), and Hamilton Depression Rating Scale (6 studies; *P*=.07).

[g] Proportion of patients (%) having received any kind of treatment for depression (ie, psychotherapy, antidepressants, or both).

### *Primary Outcomes*

The mean depression duration was 7.19 years higher (CI 95% 2.53-11.84) in iCBT trials than in face-to-face CBT trials (10.0 vs 2.8 years; *P*=.002). However, the mean proportion of patients with previous depression treatment was lower in iCBT trials (24.8% vs 42% in face-to-face trials; *P*=.04). The subgroup analyses found no evidence of differences for the remaining primary outcomes: age, gender, education, living area, depression score, history of depression, actual antidepressant medication, actual physical comorbidity, actual mental comorbidity, and study dropout.

### *Secondary Outcomes*

We found no evidence of differences between iCBT and face-to-face CBT studies for quality of life, having children, family status, and employment. Subgroup analysis for the proficiency with computers could not be performed due to insufficient studies.

## Discussion

### Principal Findings

To our knowledge, our study is the first to compare the baseline characteristics of patients with DDs included in the RCTs of eHealth and face-to-face CBT interventions. Overall, we found that the patients’ characteristics between eHealth and face-to-face RCTs were generally similar. This finding suggests that patients in both types of trials are comparable rather than different. However, patients in eHealth trials had a longer depression duration, and a lower proportion had received previous depression treatment.

eHealth psychological interventions have several advantages compared to face-to-face interventions. First, iCBT creates the opportunity to deliver psychological treatment to people without access to face-to-face therapy [12]. Thus, iCBT can help patients avoid traveling to physical consultations and can mitigate the shortage of professionals [6,34-36]. Second, guided iCBT probably represents the most economical option for the short-term treatment of adults with mild-to-moderate major depression [24]. Third, eHealth interventions have become highly automated, which enhances fidelity with treatment protocols [33,37]. Fourth, digital interventions are becoming acceptable for patients and therapists, particularly since the COVID-19 pandemic [6,24,38]. In fact, surveys suggest that iCBT guided by therapists could become a preferred option for patients over in-person CBT or medication [39].

We assumed that the baseline characteristics of patients in eHealth and face-to-face psychotherapeutic intervention RCTs would differ. For example, we hypothesized that patients in eHealth RCTs would be younger due to their familiarity with computers and frequent use of social media [40]. However, we did not find differences regarding the patients’ age, which is supported by recent literature that suggests that older adults are becoming more computer literate and that iCBT could therefore be a treatment opportunity [41]. Conversely, available research suggests that younger people are more likely to drop out of iCBT, but the reason for that result remains unknown [42]. Our study was not able to confirm this finding.

Our study found that patients in eHealth RCTs presented a longer depression duration but had received previous depression treatment in a lower number. The longer depression duration in eHealth RCTs could be explained by the fact that patients in eHealth trials perceive barriers concerning face-to-face treatments, and therefore, eHealth treatment might be more attractive to them. Conversely, we expected that patients in eHealth RCTs would present more severe depression since a lower proportion had received treatment for depression. However, our analyses did not support this assumption. Finally, our findings might indicate that eHealth trials attract patients who postpone earlier treatment attempts, but future research should be conducted to confirm this finding.

Our searches identified a high number of RCTs, which confirms the recent expansion of research into digital interventions [6,43]. However, the baseline characteristics of the trial participants were poorly and heterogeneously reported. This finding represents a major limitation of the available research, as the role of these characteristics is essential to understand under what conditions eHealth psychotherapeutic interventions will work. For example, few studies reported the patients' proficiency with computers, which is critical to explain the lack of effect of a digital intervention. Additionally, there is room for improvement in the reporting of socioeconomic characteristics, such as working conditions or family status, which are essential features to understand which type of patients benefit from eHealth interventions. Furthermore, it would be of interest to know about the comorbidities directing a patient to a certain treatment method as well as previous treatments against DDs.

This incomplete reporting highlights the need to agree to a consensus-based minimum set of baseline characteristics that should be measured and reported in all RCTs of eHealth psychotherapeutic interventions. Once the list is defined, consensus should be achieved on how to measure these characteristics, such as which measurement instruments should be selected to measure proficiency with computers. Finally, the reporting of these characteristics should be encouraged in future RCTs to allow the assessment of the applicability of the study findings.

## Strengths and Limitations

Our study had several limitations, but we tried to overcome them by following rigorous methods [44].

First, our searches may have missed eligible studies. Particularly, we limited the searches from 1990 onward as no eligible study would have been published before. The restriction until 2017 was because the searches were executed that year, and we did not have the resources to update them. However, we did not attempt to perform a systematic review and, thus, include all the studies in this field. We consider that the 58 included RCTs probably give an unbiased view of the situation in this research field.

Second, our matching process by relevant characteristics may have minimized differences between subgroups. Moreover, otherwise eligible studies were excluded because we could not find their matching pair. However, we consider that the matching process minimized confounding in the subgroup analyses (see below).

Third, subgroup analysis is a technique with considerable pitfalls. Nevertheless, we followed established guidelines to overcome the main limitations. (1) We prevented post hoc analyses and undue emphasis on particular findings by choosing the analyses in advance with clear rationale [45,46]. (2) We found a high number of studies for most outcomes, which increased the statistical power (which is usually low in subgroup analyses). For example, 6 analyses presented at least 20 studies per subgroup. (3) We did not simply compare the statistical significance of the results in each subgroup and performed formal significance tests [31,33]. Moreover, we estimated the difference in means between subgroups and its 95% CI for continuous outcomes, which allowed us to judge if the differences were clinically relevant [33]. (4) We interpreted the results cautiously. We acknowledge that subgroup comparisons are exploratory analyses that are observational by nature [31,44]. As studies are not allocated randomly to each subgroup, we cannot assume that the subgroup populations were identical except for the intervention type (internet-based or face-to-face). Thus, the results from a subgroup analysis are prone to confounding by other study-level characteristics [31]. Similarly, although we found differences in 2 characteristics between subgroups (depression duration and previous treatment for depression), we cannot conclude that they were due to the type of intervention delivery (eHealth or face-to-face).

Fourth, there is an increasing risk of type 2 error concurrent with the number of analyses, which was 16 in our case [31]. Although we did not adjust the significance level to account for multiple analyses, we assessed the impact of this decision by sensitivity analysis. After setting a new .006 threshold for statistical significance according to the proposal by Jakobsen et al [47], only the depression duration at baseline maintained its statistical significance. Thus, our general conclusion did not change; overall, patients in iCBT and face-to-face RCTs had similar sociodemographic and depression characteristics.

## Conclusions

This is the first study comparing the baseline characteristics of patients with DDs included in the RCTs of eHealth and face-to-face psychotherapeutic interventions. Overall, our study did not find differences in the patients' characteristics between eHealth and face-to-face RCTs. However, patients in eHealth trials had a longer depression duration, and a lower proportion had received previous depression treatment. This finding might indicate that eHealth trials attract patients who postpone earlier treatment attempts. Our findings highlight a need to improve the reporting of the baseline characteristics of patients included in the RCTs of eHealth psychotherapeutic interventions.

## Acknowledgments

We thank Christopher James Rose (Norwegian Institute of Public Health) and Alfonso Muriel and Borja Manuel Felix (Hospital Universitario Ramón y Cajal) for their statistical support.

## Authors' Contributions

CMW, JB, and VA defined the research question. VA performed the searches. All authors discussed the inclusion and matching criteria. Lena Kümmel and VA selected the studies. VA matched the included studies. Lena Kümmel and VA independently extracted the outcome data. The data extraction was cross-checked by JLA. JLA and VA performed the analysis. JLA and VA wrote the manuscript under the supervision of JB. All authors approved the final manuscript.

## Conflicts of Interest

CMW has active research grants to the University for digital health projects from the German health care Innovation Fund, and Newsense Lab GmbH. Board positions related to digital health for mind and body (nonpaid) are as follows: Co-Director of the Digital Society Initiative of the University of Zurich and President Fachverband Mind Body Medicine. All other authors do not have any conflict related to the content of the manuscript.

Multimedia Appendix 1
Key publications.
[DOCX File , 15 KB - jmir_v24i9e36978_app1.docx ]

Multimedia Appendix 2
Included studies.
[DOCX File , 89 KB - jmir_v24i9e36978_app2.docx ]

Multimedia Appendix 3
Summary of the included studies.
[DOCX File , 16 KB - jmir_v24i9e36978_app3.docx ]

Multimedia Appendix 4
Meta-analyses of baseline characteristics.
[DOCX File , 50 KB - jmir_v24i9e36978_app4.docx ]

## References

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
2. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1 [accessed 2022-12-14]
3. Bock JO, Luppa M, Brettschneider C, Riedel-Heller S, Bickel H, Fuchs A, et al. Impact of depression on health care utilization and costs among multimorbid patients--from the MultiCare Cohort Study. PLoS One 2014;9(3):e91973 [FREE Full text] [doi: 10.1371/journal.pone.0091973] [Medline: 24638040]
4. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED consortium. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
5. Comprehensive Mental Health Action Plan 2013-2030. World Health Organization. 2021 Sep 21. URL: https://www.who.int/publications/i/item/9789240031029 [accessed 2022-08-10]
6. Malhi GS, Bell E, Bassett D, Boyce P, Bryant R, Hazell P, et al. The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. Aust N Z J Psychiatry 2021 Jan 22;55(1):7-117. [doi: 10.1177/0004867420979353] [Medline: 33353391]
7. NICE. Depression in adults: recognition and management. National Institute for Health and Care Excellence. 2009 Oct 28. URL: https://www.nice.org.uk/guidance/cg90 [accessed 2022-08-10]
8. Swartz HA, Fournier J. Can network meta-analysis substitute for direct comparisons in psychotherapy trials? JAMA Psychiatry 2019 Jul 01;76(7):678-679. [doi: 10.1001/jamapsychiatry.2019.0243] [Medline: 30994866]
9. Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry 2020 Feb 10;19(1):92-107 [FREE Full text] [doi: 10.1002/wps.20701] [Medline: 31922679]

10. Cuijpers P, Quero S, Noma H, Ciharova M, Miguel C, Karyotaki E, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry 2021 Jun;20(2):283-293 [FREE Full text] [doi: 10.1002/wps.20860] [Medline: 34002502]
11. Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med 2013;10(5):e1001454 [FREE Full text] [doi: 10.1371/journal.pmed.1001454] [Medline: 23723742]
12. Molloy A, Ellis DM, Su L, Anderson PL. Improving acceptability and uptake behavior for internet-based cognitive-behavioral therapy. Front Digit Health 2021 Mar 25;3:653686 [FREE Full text] [doi: 10.3389/fdgth.2021.653686] [Medline: 34713125]
13. Königbauer J, Letsch J, Doebler P, Ebert D, Baumeister H. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
14. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
15. Reins JA, Buntrock C, Zimmermann J, Grund S, Harrer M, Lehr D, et al. Efficacy and moderators of internet-based interventions in adults with subthreshold depression: an individual participant data meta-analysis of randomized controlled trials. Psychother Psychosom 2021;90(2):94-106 [FREE Full text] [doi: 10.1159/000507819] [Medline: 32544912]
16. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
17. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019 Jul 01;76(7):700-707 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
18. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
19. Luo C, Sanger N, Singhal N, Pattrick K, Shams I, Shahid H, et al. A comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: a systematic review and meta-analysis. EClinicalMedicine 2020 Jul 27;24:100442 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100442] [Medline: 32775969]
20. Christ C, Schouten MJ, Blankers M, van Schaik DJ, Beekman AT, Wisman MA, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: systematic review and meta-analysis. J Med Internet Res 2020 Sep 25;22(9):e17831 [FREE Full text] [doi: 10.2196/17831] [Medline: 32673212]
21. Clark DM, Canvin L, Green J, Layard R, Pilling S, Janecka M. Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. Lancet 2018 Feb 17;391(10121):679-686 [FREE Full text] [doi: 10.1016/S0140-6736(17)32133-5] [Medline: 29224931]
22. Lindner P, Nyström MB, Hassmén P, Andersson G, Carlbring P. Who seeks ICBT for depression and how do they get there? effects of recruitment source on patient demographics and clinical characteristics. Internet Interv 2015 May;2(2):221-225. [doi: 10.1016/j.invent.2015.04.002]
23. Andersson G. Using the Internet to provide cognitive behaviour therapy. Behav Res Ther 2009 Mar;47(3):175-180. [doi: 10.1016/j.brat.2009.01.010] [Medline: 19230862]
24. Health Quality Ontario. Internet-delivered cognitive behavioural therapy for major depression and anxiety disorders: a health technology assessment. Ont Health Technol Assess Ser 2019;19(6):1-199 [FREE Full text] [Medline: 30873251]
25. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008 Dec;76(6):909-922. [doi: 10.1037/a0013075] [Medline: 19045960]
26. Cuijpers P, Karyotaki E, Ciharova M, Quero S, Pineda B, Munoz R, et al. A meta-analytic database of randomised trials on psychotherapies for depression. Open Science Framework. 2019 Jul 09. URL: https://doi.org/10.17605/OSF.IO/825C6 [accessed 2022-11-15]
27. Review Manager (RevMan). Cochrane Training. Copenhagen: The Cochrane Collaboration URL: https://training.cochrane.org/online-learning/core-software/revman [accessed 2022-08-10]
28. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14:135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
29. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. J Clin Epidemiol 2006 Jan;59(1):7-10. [doi: 10.1016/j.jclinepi.2005.06.006] [Medline: 16360555]
30. Biostat. Comprehensive meta-analysis. URL: https://www.meta-analysis.com/ [accessed 2022-08-10]
31. Deeks JJ, Higgins JPT, Altman DG. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester, United Kingdom: John Wiley & Sons; 2021.

32. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
33. Borenstein M. Common Mistakes in Meta-Analysis and How to Avoid Them. New Jersey, USA: Biostat, Inc; 2019.
34. Hoeft TJ, Fortney JC, Patel V, Unützer J. Task-sharing approaches to improve mental health care in rural and other low-resource settings: a systematic review. J Rural Health 2018 Dec;34(1):48-62 [FREE Full text] [doi: 10.1111/jrh.12229] [Medline: 28084667]
35. Fu Z, Burger H, Arjadi R, Bockting CLH. Effectiveness of digital psychological interventions for mental health problems in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Psychiatry 2020 Oct;7(10):851-864 [FREE Full text] [doi: 10.1016/S2215-0366(20)30256-X] [Medline: 32866459]
36. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
37. Wolitzky-Taylor K, Chung B, Bearman SK, Arch J, Grossman J, Fenwick K, et al. Stakeholder perceptions of the barriers to receiving and delivering exposure-based cognitive behavioral therapy for anxiety disorders in adult community mental health settings. Community Ment Health J 2019 Jan 5;55(1):83-99 [FREE Full text] [doi: 10.1007/s10597-018-0250-z] [Medline: 29508179]
38. Simon N, Robertson L, Lewis C, Roberts NP, Bethell A, Dawson S, et al. Internet-based cognitive and behavioural therapies for post-traumatic stress disorder (PTSD) in adults. Cochrane Database Syst Rev 2021 May 20;5:CD011710 [FREE Full text] [doi: 10.1002/14651858.CD011710.pub3] [Medline: 34015141]
39. Apolinário-Hagen J, Harrer M, Kählke F, Fritsche L, Salewski C, Ebert DD. Public attitudes toward guided internet-based therapies: web-based survey study. JMIR Ment Health 2018 May 15;5(2):e10735 [FREE Full text] [doi: 10.2196/10735] [Medline: 29764797]
40. Yonker LM, Zan S, Scirica CV, Jethwani K, Kinane TB. "Friending" teens: systematic review of social media in adolescent and young adult health care. J Med Internet Res 2015 Jan 05;17(1):e4 [FREE Full text] [doi: 10.2196/jmir.3692] [Medline: 25560751]
41. Silfvernagel K, Westlinder A, Andersson S, Bergman K, Diaz Hernandez R, Fallhagen L, et al. Individually tailored internet-based cognitive behaviour therapy for older adults with anxiety and depression: a randomised controlled trial. Cogn Behav Ther 2018 Jul;47(4):286-300. [doi: 10.1080/16506073.2017.1388276] [Medline: 29068266]
42. Edmonds M, Hadjistavropoulos HD, Schneider LH, Dear BF, Titov N. Who benefits most from therapist-assisted internet-delivered cognitive behaviour therapy in clinical practice? predictors of symptom change and dropout. J Anxiety Disord 2018 Mar;54:24-32 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.003] [Medline: 29421369]
43. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
44. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? updating criteria to evaluate the credibility of subgroup analyses. BMJ 2010 Mar 30;340(mar30 3):c117. [doi: 10.1136/bmj.c117] [Medline: 20354011]
45. Paquette M, Alotaibi AM, Nieuwlaat R, Santesso N, Mbuagbaw L. A meta-epidemiological study of subgroup analyses in cochrane systematic reviews of atrial fibrillation. Syst Rev 2019 Oct 25;8(1):241 [FREE Full text] [doi: 10.1186/s13643-019-1152-z] [Medline: 31653275]
46. Gagnier JJ, Moher D, Boon H, Beyene J, Bombardier C. Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. BMC Med Res Methodol 2012 Jul 30;12:111 [FREE Full text] [doi: 10.1186/1471-2288-12-111] [Medline: 22846171]
47. Jakobsen JC, Wetterslev J, Winkel P, Lange T, Gluud C. Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. BMC Med Res Methodol 2014 Nov 21;14:120 [FREE Full text] [doi: 10.1186/1471-2288-14-120] [Medline: 25416419]

## Abbreviations

**CBT:** cognitive behavioral therapy
**DD:** depressive disorder
**iCBT:** internet cognitive behavioral therapy
**RCT:** randomized controlled trial
**SMD:** standardized mean difference

*Edited by G Eysenbach; submitted 02.02.22; peer-reviewed by E Smoktunowicz, C Salewski; comments to author 26.02.22; revised version received 16.05.22; accepted 28.07.22; published 14.09.22.*

*Please cite as:*
*Aemissegger V, Lopez-Alcalde J, Witt CM, Barth J*
*Comparability of Patients in Trials of eHealth and Face-to-Face Psychotherapeutic Interventions for Depression: Meta-synthesis*
*J Med Internet Res 2022;24(9):e36978*
*URL: https://www.jmir.org/2022/9/e36978*
*doi:10.2196/36978*
*PMID:36103217*

Review

# The Impact of Electronic Health Record Interoperability on Safety and Quality of Care in High-Income Countries: Systematic Review

Edmond Li[1], MB BCh BAO, MPH; Jonathan Clarke[1], MA MB BChir, MPH, PhD; Hutan Ashrafian[1], MBBS, MBA, PhD; Ara Darzi[1], OM, KBE, PC; Ana Luisa Neves[1], MBChB, MSc, PhD

National Institute for Health and Care Research (NIHR) Imperial Patient Safety Translational Research Centre, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

**Corresponding Author:**
Edmond Li, MB BCh BAO, MPH
National Institute for Health and Care Research (NIHR) Imperial Patient Safety Translational Research Centre
Institute of Global Health Innovation
Imperial College London
Room 1035/7, QEQM Wing
St Mary’s Hospital Campus
London, W2 1NY
United Kingdom
Phone: 44 02033127259
Fax: 44 02033126309
Email: edmond.li19@imperial.ac.uk

## Abstract

**Background:** Electronic health records (EHRs) and poor system interoperability are well-known issues in the use of health information technologies in most high-income countries worldwide. Despite the abundance of literature exploring their relationship, their practical implications on patient safety and quality of care remain unclear.

**Objective:** This study aimed to examine how EHR interoperability affects patient safety, or other dimensions of care quality, in high-income health care settings.

**Methods:** A systematic search was conducted using 4 web-based medical journal repositories and grey literature sources. The publications included were published in English between 2010 and 2022, pertaining to EHR use, interoperability, and patient safety or care quality in high-income settings. Screening was completed by 3 researchers in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Risk of bias assessments were performed using the Risk of Bias in Non-randomized Studies of Interventions and the Cochrane Risk of Bias 2 tools. The findings were presented as a narrative synthesis and mapped based on the Institute of Medicine’s framework for health care quality.

**Results:** A total of 12 studies met the inclusion criteria to be included in our review. The findings were categorized into 6 common outcome measure categories: patient safety events, medication safety, data accuracy and errors, care effectiveness, productivity, and cost savings. EHR interoperability positively influenced medication safety, reduced patient safety events, and reduced costs. Improvements in time saving and clinical workflow are mixed. However, true measures of effect are difficult to determine with certainty because of the heterogeneity in the outcome measures used and notable variation in study quality.

**Conclusions:** The benefits of EHR interoperability on the quality and safety of care remain unclear and reflect extensive heterogeneity in the interventions, designs, and outcome measures used. The establishment of common health information technology research outcome measures would support higher-quality research on the topic. Future research efforts should focus on both the positive and negative impacts of interoperable EHR interventions and explore patient perspectives, given the growing trend for patient involvement and stewardship over their own electronic clinical data.

**Trial Registration:** PROSPERO CRD42020209285; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=209285

**International Registered Report Identifier (IRRID):** RR2-10.1136/bmjopen-2020-044941

***(J Med Internet Res 2022;24(9):e38144)*** doi:10.2196/38144

**KEYWORDS**

electronic health records; interoperability; patient safety; systematic literature review; health information exchange; digital health

# Introduction

## Background

Electronic health records (EHRs) have become a mainstay digital solution for high-income health systems globally [1-3]. Despite the growing integration of such technologies into the routine workflows of health care providers, sizable challenges remain that prevent EHRs from fulfilling their full potential. One such hurdle is the lack of interoperability.

The Healthcare Information and Management Systems Society defines interoperability as "the ability of different information systems, devices and applications (systems) to access, exchange, integrate and cooperatively use data in a coordinated manner, within and across organizational, regional and national boundaries, to provide timely and seamless portability of information and optimize the health of individuals and populations globally" [4]. However, this may have drastically different implications depending on one's role and perspective in a health system.

From a technological standpoint, EHR interoperability can be defined as "the ability of two or more applications to communicate effectively without compromising the content of the transmitted EHR" [5]. However, barriers such as hardware, syntax, and system usability often hinder the implementation of this vision [6-8]. The adoption of common standards in terminology, content, and security has been proposed to facilitate different levels of interoperability within or across health care settings [4,9].

From a public health, administrative, or policy-making perspective, EHR interoperability may entail "electronic health information that is shared appropriately between healthcare and public health partners in the right format, through the right channel at the right time" [10]. For end users such as health care providers and patients, the notion of interoperability often focuses more on the practical functionalities of EHR interoperability. For health care providers, this may include being able to remotely access care records from another health care setting, electronically correspond with other providers, and coordinate complex care plans with external health care organizations [2,3,11-13]. For patients, interoperability may mean a more seamless experience when seeking care from various health care providers or feeling an increased sense of empowerment by having greater access to their health records [14].

Lack of interoperability may negatively impact all 6 dimensions of quality of care outlined by the Institute of Medicine (IoM) framework: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity [15-17]. These shortcomings may range from inaccurate or fragmented patient health records across multiple care providers and delayed communication between care teams to increased costs resulting from duplicated efforts by staff and health care resources used [7,18-20]. The introduction of EHRs has been attributed to benefits such as improving health care worker productivity, facilitating public health disease surveillance and research, and even generating cost savings [21]. While the implementation of EHR interoperability was speculated to realize further benefits, such as improved care coordination and additional potential savings, current evidence remains mixed [3,8,22].

Although the issues associated with EHR interoperability are well recognized in high-income health care settings [8], evidence exploring its impact on the 6 dimensions of care quality remains relatively scarce. Nearly 2 decades after the initial adoption of EHR in high-income countries (HICs) such as the United States and the United Kingdom [8,22-24], there is a need to revisit and review the currently available literature to evaluate EHR interoperability, its impact on quality of care, and particularly in terms of patient safety.

## Aims

This systematic review aims to evaluate the impact of EHR interoperability on the IoM's 6 domains of health care quality in HICs [15].

# Methods

## Search Strategy

A literature search was conducted for publications published between 2010 and 2020 on 4 databases (PubMed, MEDLINE, Embase, and PsycINFO). Publications from grey literature sources and relevant papers identified from the references of the screened articles were also included. A more thorough description of the search strategy and inclusion criteria was previously published as a study protocol [25].

We performed an additional search for publications published between March 2020 and June 2022. As the onset of COVID-19 has had profound implications on many aspects of health care technologies and policies, this supplementary search was conducted to account for any new studies published during that period. Otherwise, there were no further deviations from the methods described in the previously published protocol.

## Study Selection Criteria

This systematic review included studies fulfilling the following criteria: (1) studies took place in HICs as defined by the World Bank where "the gross national income (GNI) *per capita* is higher than $13,205 USD" [26,27], (2) investigated EHRs or other health information technologies (HITs) that facilitate the sharing of clinical information between health care providers, and (3) contained outcomes concerning patient safety or quality of care. Only the studies published in English were included. No other filters, such as the study design, type, or publication country of origin, were used. Screening and selection of publications were performed by a total of 3 reviewers. Two reviewers initially independently screened the body of articles derived from the database searches to be considered for inclusion. This was performed iteratively at the title, abstract, and full-text levels in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram [28]. Cohen κ statistic was used to measure interrater reliability [29]. Discrepancies in article selection were arbitrated by a third reviewer.

## Data Extraction

Data extraction was completed by the first reviewer using a standardized Microsoft Excel spreadsheet. The content was then reviewed by the other 2 reviewers to ensure data quality and consistency. The characteristics and data extracted from each study included the name of the authors, year of publication, study design, study setting, study population and size, outcome measures, and general findings.

## Risk of Bias and Quality Assessment

The Cochrane Risk of Bias Tool was used to assess randomized control trials for bias, whereas the "Risk of Bias in Non-Randomized Studies—of Interventions" tool was used for nonrandomized trials [30]. Risk of bias assessments were conducted by 2 reviewers and any disagreements were resolved by a third investigator.

## Data Synthesis

A narrative synthesis of the findings was conducted. Relevant findings and outcome measures were grouped into subcategories and organized based on the 6 domains found in the IoM health care quality framework. Given the variety of outcome measures used in the included studies, a meta-analysis was not performed.

# Results

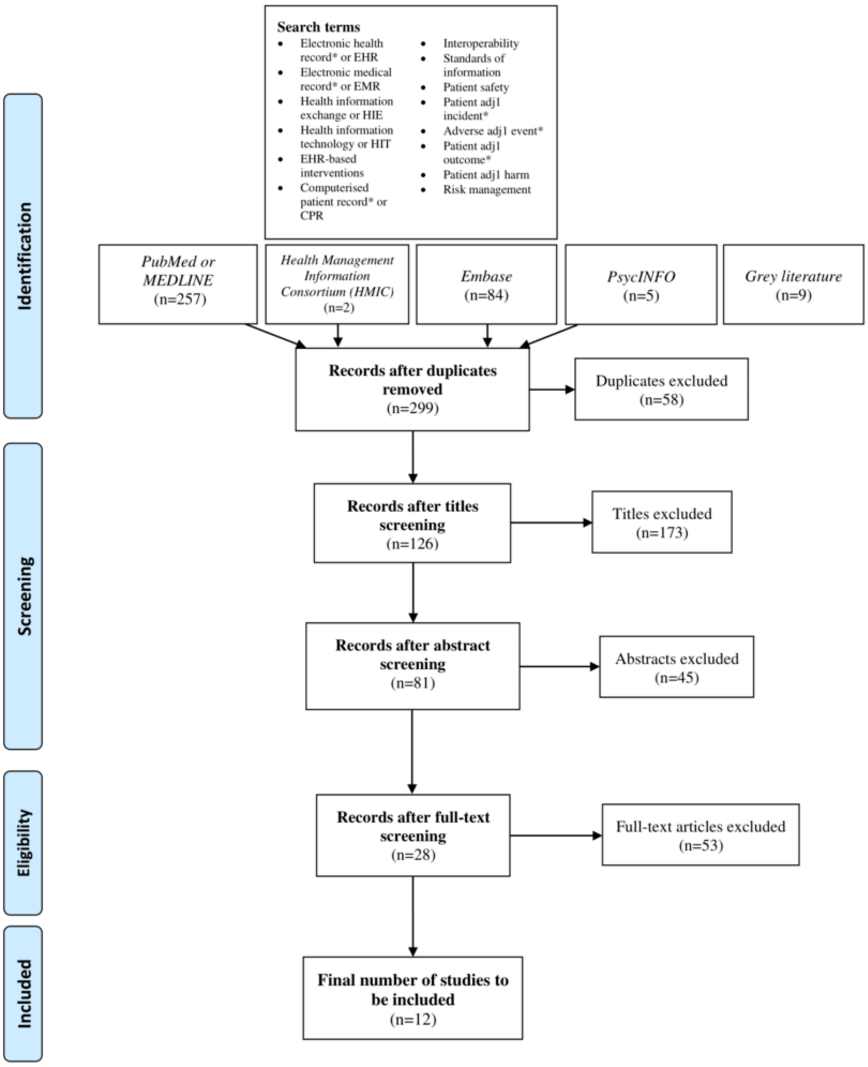
## Overview

The initial search using computerized databases yielded 299 publications (Figure 1). After screening the titles, 173 publications were excluded as they did not meet the inclusion criteria. Upon further screening of the abstracts, 45 articles were rejected because they were either not relevant or were not studies, but rather commentaries or opinion pieces. Following full-text screening and agreement among the 3 reviewers, 53 publications were excluded because they did not satisfy all the requirements of the PICO (population, intervention, control, and outcomes) inclusion criteria.

A final total of 12 papers were selected for this systematic review.

κ statistic for full-text screening was 0.52, indicating fair agreement [29].

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) systematic review search strategy and screening process flow diagram. Search terms appear as used in Li et al [25].



### Description of Included Studies

The 12 selected studies were published between 2013 and 2020. The included studies were predominantly nonrandomized trials, including observational studies (n=5) [31-35], cross-sectional studies (n=2) [36,37], retrospective analyses of patient safety reports (n=2) [38,39], case studies (n=1) [40], and a simulation study (n=1) [41]. Only one randomized controlled trial was included [42]. Further details on the included studies have been provided in Table 1.

**Table 1.** Summary characteristics of included studies.

| Study | Publication year | Journal | Study type | Stated aims or objectives | Date or duration of intervention | Study population or settings |
|---|---|---|---|---|---|---|
| Reed et al [34] | 2020 | *American Journal of Managed Care* | Observational study | • To examine whether providers' timely access to clinical information through shared inpatient-outpatient EHRsa was associated with follow-up visits, return emergency department visits, or readmissions after hospital discharge in patients with diabetes. | 2005-2011 | 241,510 hospitalized patients with diabetes discharged home from 17 hospitals where a new inpatient EHR system is being gradually introduced which integrates with an existing outpatient EHR system. |
| Wong et al [35] | 2020 | *Journal of the American Medical Directors Association* | Observational study | • To assess the impact of implementing a new electronic medical records transfer mechanisms or process to improve the transfer of medical records when transitioning patients between nursing facilities and acute settings | 2020 | HOPE[b] SNF[c] Collaborative of 25 nursing facilities working with 3 hospitals in a local health network. |
| Howe et al [38] | 2018 | *JAMA* | Retrospective analysis of patient safety reports | • To explore how EHR usability can contribute to patient harm by reviewing patient safety reports from the Pennsylvania Patient Safety Authority database. | 2013-2016 | Patient safety reports from the Pennsylvania Patient Safety Authority database derived from 571 health care facilities. |
| Biltoft et al [40] | 2018 | *American Journal of Health-System Pharmacy* | Case study | • To improve IV[d] infusion:<br>• medication safety<br>• accuracy, timeliness, and efficiency of IV medication documentation<br>• Free up pharmacist and nurse time for direct patient care<br>• Increase revenue by improving reimbursement for IV medications in outpatient areas | October 2013, lasting for 7 months | Regional health system consisting of 8 hospitals, excludes NICUs[e] |
| D'Amore et al [36] | 2018 | *AMIA[f]symposium* | Cross-sectional study | • To examine testing artifacts from recent certification through automated tooling and manual review to identify compatibility and usability issues. | January 2018 | 854 C-CDA[g] documents were selected from the Office of the National Coordinator for Health Information Technology publicly available repository. After screening for duplicates, invalid XML, and documents not confirming to C-CDA 2.1 standards, 401 C-CDA documents were examined |
| Adams et al [39] | 2017 | *Applied Clinical Informatics* | Retrospective analysis of patient safety reports | • Overall study was to understand patient safety consequences resultant from interoperability issues between EHRs and HIT[h]. Specific objectives were:<br>• To identify patient safety incident reports that reflect EHR interoperability challenges with other health IT.<br>• To perform a detailed analysis of these reports to understand the health IT systems involved, the clinical care processes impacted, whether the incident occurred within or between provider organizations, and the reported severity of the patient safety events. | 2009-2016 | 1.735 million PSE[i] reports from the Pennsylvania Patient Safety Authority's Pennsylvania Patient Safety Reporting System, attained through the ISMP[j], and a large health care system in the Mid-Atlantic United States; 209 (8%) PSE reports of the 2625 health IT reports were determined to be related to interoperability between the EHR and another health IT system. |

| Study | Publication year | Journal | Study type | Stated aims or objectives | Date or duration of intervention | Study population or settings |
|---|---|---|---|---|---|---|
| Elysee et al [31] | 2017 | *Medicine (United States)* | Observational study | - To empirically examine how the 3 capabilities (HIE[k], interoperability, medication reconciliation) influence one another so the appropriate policy can be applied where it can have the greatest impact. | 2013 | AHA[l] Annual IT Survey responses; 1330 hospitals were included. 2013 AHA Annual Survey IT Supplement database to obtain a nationally representative sample of nonfederal acute care hospitals that (1) include acute care general medical and surgical, general children's, and cancer hospitals (2) use any type of electronic exchange or sharing of care summaries with other providers |
| Motulsky et al [32] | 2016 | *Studies in Health Technology and Informatics* | Observational study | - Evaluated the accuracy and usability of SQIM software for documenting the list of current medications for patients at admission to hospital and comparing with medication lists with pharmacies via fax. | June 2014 to January 2015 | 111 patients, average age of 76 years, 51% female, average of 11 medications. On the basis of tertiary care center in Montreal, Canada |
| Akbarov et al [37] | 2015 | *Drug Safety* | Cross-sectional study | - To investigate the feasibility of linked primary and secondary care EHR data for surveillance of medication safety. Objectives included assessing the prevalence of 22 medication safety indicators, investigating associations with patient and practice characteristics, and investigating variation between general practices. | April 2012 | 52 general practices affiliated with 205,519 patients in Salford, United Kingdom |
| Munck et al [42] | 2014 | *Danish Medical Journal* | Randomized control trial+Likert scale questionnaire | - Examines time expenditure and impact on workflow the use of an integrated shared medical record has on medication reconciliation at hospital admissions | June 2010 | Sixty-two patient consultations, 18 physicians participated from the accident and emergency department at Køge Hospital—a university-affiliated hospital. |
| Koldby et al [41] | 2013 | *Studies in Health Technology and Informatics* | Simulation study | - To evaluate how integration between digital dictation and EHRs impacts workflow, and functionality, and identify areas requiring further improvement. | N/A[m] | Three doctors (2 surgeons, one pediatrician) and 3 medical secretaries, Herlev Hospital in Copenhagen, Denmark |
| Lee et al [33] | 2013 | *Studies in Health Technology and Informatics* | Observational study | - To develop and implement a workflow-based multidisciplinary hand-over information system, integrated with medical record browsing, multidisciplinary hand-over, and event tracking to improve the correctness and effectiveness of communication among the medical team members. | 2 years, auditing was completed every 3 months | 40+ seed anchors were trained on the use of the cross-disciplinary team hand-over information system. They were responsible for training nurses in their respective wards; no further detail on sample size |

---

[a] EHR: electronic health record.

[b] HOPE: Health Optimization for Elders.

[c] SNF: Skilled Nursing Facility.

[d] IV: intravenous.

[e] NICU: neonatal intensive care unit.

[f] AMIA: American Medical Informatics Association.

[g] C-CDA: consolidated clinical document architecture.

[h] HIT: health information technology.

[i] PSE: patient safety event.

[j] ISMP: Institute for Safe Medication Practices.

[k] HIE: health information exchange.

[l] AHA: American Hospital Association.

[m] N/A: not applicable.

## Description of Interventions in Included Studies

The intervention of interest examined in this systematic review is the implementation of interventions that intend to improve EHR interoperability with other EHRs or health IT systems. The review incorporated interventions that aimed to reduce interoperability errors, inaccurate patient records with inappropriate units of measurement, incorrect medication doses, and omission of codes or units of measure in laboratory results [36,43].

Other studies included in this review encompassed interventions that enabled interoperability between hospital-based EHRs, primary care databases, as well as medical devices such as infusion pumps [40]. Across the reviewed studies, interoperability ranged from bidirectional reading and writing of data to unidirectional writing of clinical information to a medical device or record.

## Outcomes

### *Summary of Outcome Measures*

The outcomes explored by the 12 studies varied considerably and could be broadly grouped into 6 categories belonging to 3 main domains found in the IoM health care quality framework: (1) patient safety events (PSEs); (2) medication safety; (3) data sharing, accuracy, and errors; (4) care effectiveness; (5) productivity; and (6) cost savings. As all outcome subtypes could be categorized into 3 out of the 6 domains (ie, safety, effectiveness, and efficiency), only these 3 domains are included in our review for clarity. It should be noted that medication safety typically would be classified as a part of patient safety. However, as it is a common outcome specifically evaluated across many of the included studies, it is considered an independent outcome of interest in this review. Table 2 provides an overview of the outcome measures featured in each study. A detailed summary of the main findings is presented in Multimedia Appendix 1 [31-42].

**Table 2.** Outcome measures explored by the included studies, mapped onto the Institute of Medicine health care quality framework [15].

| | Safety | | | Effectiveness | Efficiency | |
|---|---|---|---|---|---|---|
| Study, year | Patient safety events | Medication safety | Data sharing, accuracy, and errors | Care effectiveness | Productivity | Cost savings |
| Reed et al [34], 2020 | ✓[a] | | | ✓ | | |
| Wong et al [35], 2020 | | | ✓ | | | |
| Howe et al [38], 2018 | ✓ | ✓ | | | | |
| Biltoft et al [40], 2018 | ✓ | ✓ | ✓ | | | ✓ |
| D’Amore et al [36], 2018 | | ✓ | ✓ | | | |
| Adams et al [39], 2017 | ✓ | ✓ | | | | |
| Elysee et al [31], 2017 | | | ✓ | | | |
| Motulsky et al [32], 2016 | | ✓ | ✓ | | | |
| Akbarov et al [37], 2015 | | ✓ | | | | |
| Munck et al [42], 2016 | | ✓ | | | ✓ | |
| Koldby et al [41], 2013 | ✓ | | | | ✓ | |
| Lee et al [33], 2013 | ✓ | | | | ✓ | |

[a] ✓: denotes that the specified outcome measure was present and explored in the study.

### *Patient Safety Events*

A total of 6 studies included PSEs as outcome measures [33,38-41]. In 2 studies that reviewed patient records, EHR interoperability was responsible for only a small minority of safety events. Howe et al [38] found that 18.1% (102/557) of EHR-related PSEs were specifically attributed to interoperability issues. Similar results were reported by Adams et al [39], where 7.9% (209/2625) of patient safety incidents were related to problems with EHR interoperability. Notably, however, most of the problems that resulted in safety events that reached patients did not cause any direct harm (111/209, 53.1%) [39]. EHR interoperability issues resulting in medication (42/209, 20%), laboratory (33/209, 15.7%), or radiology-related (22/209, 10.5%) events comprised the largest categories of safety incidents identified [39]. These PSEs were more common when sharing clinical information among different EHR systems within a health care facility rather than when communicating with other health care providers externally [39].

The study by Reed et al [34] also examined the rates of adverse clinical events in patients with diabetes 30 days after discharge from the hospital using a new shared, integrated EHR. Adverse clinical events were determined using emergency department visits and hospital readmissions as proxy measures. However, the study found no significant change in the rates of emergency department visits (16.4% vs 16.7%) or readmissions (9.5% vs 9.4%) [34].

### *Medication Safety*

Medication-related measures are some of the most used outcome measures for assessing the impact of EHR interoperability on patient safety.

For tasks such as medication reconciliation, Munck et al [42] evaluated the impact of emergency department clinicians using an EHR system interoperable with national shared medication records, compared with a standalone EHR system. Participating clinicians reported "unambiguous support" of shared medication record integration due to their perceived utility and ease with which it can be incorporated with little detriment to their workflow [42]. The clinician workload was not perceived to be different between interoperable EHRs and standalone systems [42]. In contrast, improvements in the accuracy of medication lists and communication between health care providers and patients have not been observed [42].

An observational study by Motulsky et al [32] investigated the accuracy of a new digital application introduced for documenting medications when patients present to the hospital by integrating data derived from various points along a patient's encounter with the health care system from prescription to medication review. This was compared against a nonintegrated list from a community pharmacy. Approximately 64% of patients had discrepancies in their hospital's medication lists, which were categorized into 3 main types of errors: (1) false positive, that is, medications listed that should not have been present, (2) false negative, that is, medications not found on the list but should have been present, and (3) duplication of medications. Of the 111 participants, 442 discrepancies were reported in their medication lists: (1) 44.6% had medications on the hospital-based list that should not have been, (2) 43.9% had current medications missing from hospital-based lists that should have been present, and (3) 11.5% contained duplicates [32].

Findings from Howe et al [38] highlighted that medication-administration safety events accounted for 37% (n=207) of the 557 EHR-related patient safety reports of events that reached patients [38]. This categorization included adverse drug events and incorrect medication dosing or route of administration [38].

Interoperability between EHR systems and other medical devices, such as infusion pumps, has also been found to be beneficial for patient safety. In the study by Biltoft et al [40], the authors noted that the implementation of an interoperable smart pump-EHR program resulted in an average number of alerts that reportedly decreased by 22% (n=1845 vs n=1447) monthly. The corresponding number of infusions requiring intervention by health care staff also dropped by nearly 20% (n=119 vs n=96), in addition to an annual reduction of staff-reported safety events from four to one [40]. By using smart pumps to prepopulate infusion parameters based on clinical data retrieved directly from interoperable EHRs, approximately 3.5 million data entry keystrokes and opportunities for errors across 8 participating hospitals were avoided monthly [40]. Clinicians have also identified an additional benefit of being able to adjust intravenous medication administration in response to a patient's changing clinical parameters accessible directly from the EHR [40].

Finally, Elysee et al [31] examined the relationship between hospitals implementing health information exchanges (HIEs), interoperability, and medication reconciliation [31]. For successful adoption and use of HIEs, clinical information-sharing functionality is not only needed between various secondary or community-based health care facilities but also with patients. The authors concluded that these 3 capabilities are closely linked and that stalling the implementation of one of these elements in hospitals would have a detrimental impact on the adoption of the other two.

### *Data Sharing, Accuracy, and Errors*

The relationship between EHR interoperability and its impact on data accuracy, sharing, and errors was explored in 5 studies.

Taking a more longitudinal perspective, D'Amore et al [36] investigated how data quality in EHRs changed in American Veterans Affairs (VA) hospitals through the increasing use of systems with greater interoperability. Using 3 independent evaluation tools, the authors found a general increase in the scope and accuracy of the clinical data being shared [36]. With HL7 Schematron testing, 86.3% (346/401) of electronic health documents contained 1695 errors, averaging 4.9 errors per clinical document [36]. Finally, using a data quality algorithm, 21,304 alerts (indicating issues in either the completeness or syntax of the record) were generated from the 401 documents examined, averaging 53.1 alerts per document [36]. 57% of these alerts were triggered due to issues surrounding data completeness, and 43% from syntax [36].

Compared with prior research, D'Amore et al [44] highlighted that the federal program for HIT certification has resulted in notable developments in the scope of information included in consolidated clinical document architecture documentation [44]. An example provided was that of implanted devices, a category not previously included in prior consolidated clinical document architecture versions [36,44].

Biltoft et al [40] reported that the rates of appropriate or correct patient ID entries across 8 participating hospitals increased from 35.5% to 81% as a result of the information being automatically filled in from interoperable EHRs when compared with the clinicians manually entering patient details [40]. Similarly, the authors also indicated a 22% (1845-1447) reduction in the average monthly infusion pump alerts as well as a 19% (119-96) reduction in the number of errors that necessitated reprogramming of the infusion pump [40]. Finally, the authors also observed a 33% (166-111) decline in the mean number of cancelled infusions per month with the introduction of EHRs interoperable with smart pumps for medication infusion [40].

A more recent study by Wong et al [35], performed during the initial months of the COVID-19 pandemic, explored the introduction of a novel workflow process using software to integrate outpatient e-fax data inputs to inpatient EHRs to improve data sharing between transitions of care and limit disease transmission. Nine weeks after debuting their new process, the authors reported that it was utilized in 287 instances across the three-hospital system, with uptake and use trending positively [35]. Feedback from hospital staff was also largely positive, with only minor instances of data entry errors such as the system not handling double-sided documents properly and delays taking more than the usual hour [35]. However, the authors provided little detail as to why the number of e-faxes received, and the timeframe of nine weeks post-implementation was selected as the outcome measure for assessing the effectiveness of their intervention. As such, their purported findings must be interpreted with circumspection, given the limitations of the study design and likely biases present.

### *Care Effectiveness*

One study examined how the use of interoperable EHRs by clinicians across inpatient and outpatient settings affected health outcomes and the follow-up care that patients received. Reed et al [34] explored how the rates and modes of follow-up for patients with diabetes changed after the incremental introduction of a new inpatient EHR system integrated with an existing outpatient EHR. The authors found a statistically significant reduction in the rates of in-person office visits (56%-50%) and outpatient laboratory testing (32%-31%) [34]. However, secure messaging and phone calls remain unchanged. Overall, the follow-up rates decreased from 73% to 69%.

### *Productivity*

Of the 12 studies, 3 investigated the impact of interoperability on the efficiency of clinicians [33,41,42]. Of these, one study quantified the exact time saved by the clinicians [42]. Munck et al [15] primarily assessed the time spent by clinicians to perform medication reconciliation. They found that the time expended per patient using an EHR interoperable with patients' historical medication records vs. a standalone EHR system was not significantly lengthened or statistically significant (5 minutes 27 seconds vs 4 minutes 15 seconds) [15].

Two studies captured the perceived time savings from health care providers [33,41]. Lee et al [33] described the introduction of a multidisciplinary hand-over information system that is interoperable with medical records and event tracking. Compared with an existing paper-based Kardex system, nurses reported a 50% time saving [33]. It should be noted, however, that these findings should be interpreted with caution given the limited details provided and the high risk of bias.

A simulation study conducted by Koldby et al [41] primarily examined whether integrating a digital dictation system into an EHR system would reduce unintended clinical incidents. While the authors expected improvements in this domain, they also hypothesized the simultaneous emergence of some unintended consequences, such as clinical documentation storage errors resulting from the novel intervention being introduced [41]. The participants reported notable improvements in their clinical workflows and time savings, as access to dictation services, transcripts, and medical records can be made via one click in the EHR itself. However, this benefit is ultimately offset by the limited functionality of the EHR system owing to the suboptimal integration of the dictation system [41]. Frequent system lockups, inability to open windows to other commonly used applications when dictations were being performed, and poor interoperability with other hospital systems (eg, retrieving data from laboratory information systems) all contribute to curtailing potential workflow benefits [41]. Furthermore, no workflow-related benefits were observed for other supporting clinical staff, such as medical secretaries [41].

### *Cost Savings*

Of the 12 studies reviewed, only Biltoft et al [40] explored the cost savings made possible by using interoperable EHR systems. Their study attributed the introduction of smart pump-EHR interoperability to a reduction in lost revenues (US $980,000 vs US $610,000) [40]. Other indirect cost benefits, such as reducing documentation times by nurses resulting in annual cost savings of US$ 2,452,800 were also hypothesized [40].

## Risk of Bias Assessment

Upon assessing the overall risk of bias in the included studies (n=12), 7 were determined to be of low risk, 1 of moderate risk, 3 of serious risk, and 1 of critical risk (Figures 2-4). A study was considered "overall low risk" if at least 50% of the domains were rated "low risk." Studies that had 2 or more domains rated "moderate risk" or higher were rated based on the most numerous lowest-risk domain rating. Despite the heterogeneity of previous studies, many of the studies' data were derived from examining commonly available parameters in EHRs themselves (eg, medication lists, PSEs, or alerts) or subjective surveys of health care workers before and after changes to improve EHR interoperability were made. No studies provided evidence of a preceding registered study protocol being published. Incomplete or missing data appeared to be the most common risk of bias; many of these studies only presented their aggregated findings, rather than sharing a more detailed quantitative breakdown of their assembled data.

**Figure 2.** Risk of Bias in Non-Randomized Studies—of Interventions traffic lights plot for domain-level risk of bias judgments for nonrandomized studies.

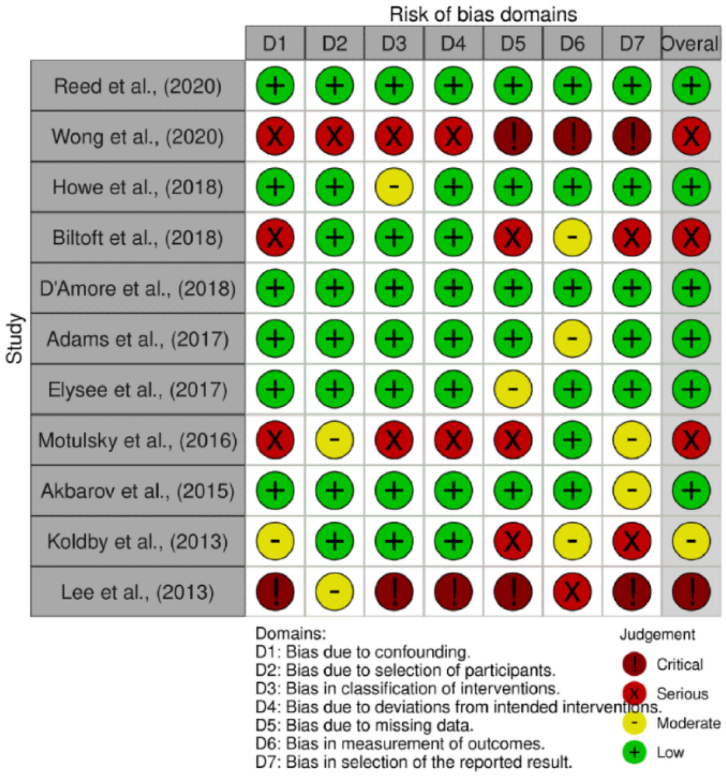


**Figure 3.** Risk of Bias in Non-Randomized Studies—of Interventions summary plot of biases present in nonrandomized studies included in the review [31-42].

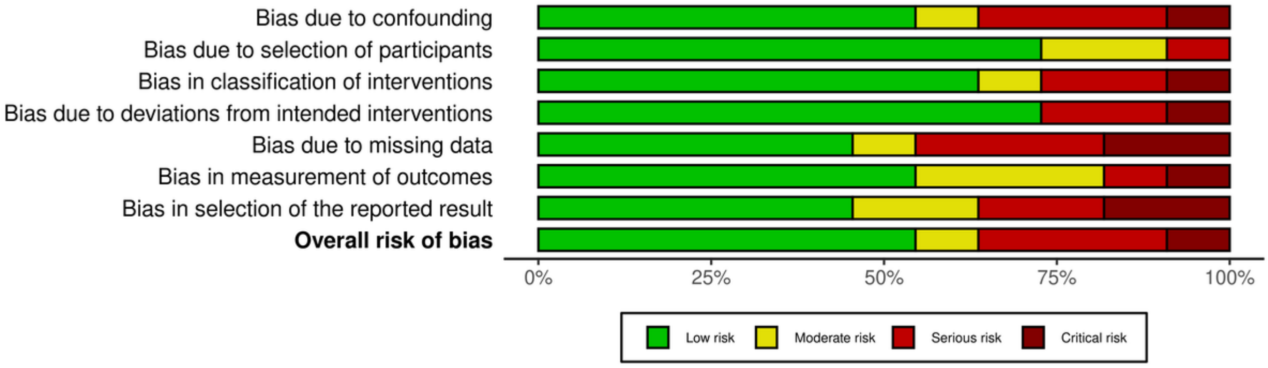
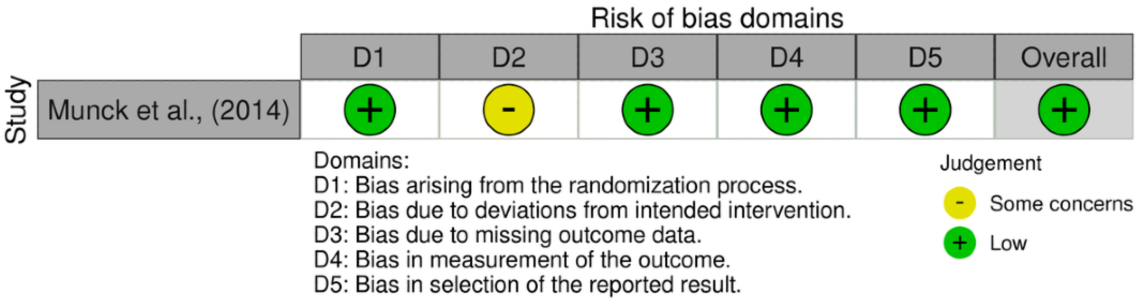


**Figure 4.** Risk of Bias 2 traffic lights plot for domain-level risk of bias judgments in randomized studies.



## *Discussion*

### Summary of Principal Findings

Despite similar overall aims, study objectives, and outcome measures, the studies included in this review varied widely (Multimedia Appendix 1). Most of the studies explored more than one outcome measure (n=9). The most frequently evaluated outcome measure was medication safety or reconciliation (n=7) [32,36-40,42], followed by reported PSEs (n=6) [33,34,36,38,40,41], data sharing, accuracy, and errors (n=5)

[31,32,35,36,40], productivity (n=3) [33,41,42], care effectiveness (n=1) [34], and costs (n=1) [40].

Altogether, EHR interoperability appeared to have some beneficial effects on both PSEs and medication safety or reconciliation. Implementing interoperable EHRs was also associated with reduced data entry errors and improved overall data quality and scope [36,40]. However, the impact of EHR interoperability on timeliness and improvements to clinical workflow remains inconclusive; as described by Munck et al [42], where the time expended to complete the clinical task increased following EHR interoperability with a community medication list.

The least used outcome measure in the reviewed studies was efficiency, with only one study explicitly mentioning changes to the cost of care associated with EHR interoperability [36]. Through the introduction of an EHR system interoperable with smart infusion pumps, participating hospitals were reported to theoretically benefit from cost savings derived from lost charges for infusions in outpatient settings and lost revenue [40].

## Strengths and Limitations

This systematic review has several strengths. It evaluated a comprehensive body of evidence published between 2010 and 2022 and retrieved 12 studies. To ensure the transparency of our search strategies, a preceding protocol paper was published, and best practice guidelines were adhered to generate these results [25]. Only studies that focused specifically on EHRs, interoperability, quality of care, and patient safety were included. Two researchers were involved in the screening and reviewing processes, with a third senior researcher to arbitrate any discrepancies that arose. The final list of included studies was determined only after a consensus was reached by all 3 researchers upon iterative consultation. Our findings were subsequently mapped onto a well-established framework for care quality commonly used in health care research [15].

However, there are notable limitations to our systematic review. The first is the decision to restrict the review to include only publications completed in English and focused on HICs. This limits the ability of this review to capture the experiences of EHRs in non-English health care settings or low- and middle-income countries. Although EHR systems are commonly found in HICs, and thus would likely have more extensive experience regarding the problem of interoperability, the authors acknowledge that these inclusion criteria can introduce a risk for bias.

Second, our study highlights the positive impact of interoperable EHRs on various outcome measures; we did not identify many negative findings. This relative lack of reported negative impact can likely be attributable to publication bias, where studies with negative findings are less likely to be published, and thus not available to be included in our review. Health IT interventions with positive outcomes also often garner more robust follow-up research efforts and deflect resources away from exploring interventions that demonstrate negative results, thus further reducing the visibility of studies investigating the latter [45].

Third, the heterogeneity of study types, contextual parameters regarding the EHR interventions examined, and study outcomes made it difficult to draw any direct comparisons between the findings of the studies. Considering this heterogeneity, we have summarized these differences in detail in Tables 1 and 2 to provide a transparent overview of the differences across studies.

Finally, the studies included in our review had an overall high risk of bias. Even more robust studies that manage to provide greater specificity in their methods and results are often limited in scope or outcomes assessed. We examined these in detail using recognized tools and provided a comprehensive summary, so that the results can be interpreted despite these limitations. Together, these limitations made it challenging to draw any definitive conclusions regarding the overall magnitude of the effect that EHR interoperability has on improving patient safety or care quality, as well as its generalizability to other health systems.

## Comparison With Prior Work

Our review findings are generally in line with those of previously published systematic reviews. These earlier reviews tended to focus on the initial adoption of EHRs in acute health care settings [46,47]. As EHR availability matured, more recent reviews began to consider EHRs' relationships with other topics such as interoperability, examining how it affects outcome measures in diverse settings [45]. No systematic review has specifically investigated the relationship between interoperability and patient safety or quality of care.

A systematic review by Chaudhry et al [46] explored the effects of HIT adoption on quality of care and efficiency [46]. This study concluded that the introduction of HITs resulted in greater adherence to guidelines or protocols and had a positive effect on improving medication safety (ie, higher rates of adverse drug event identification and reduction in adverse drug events) [46]. Chaudhry et al [46] also described some negative consequences of their adoption and use, including mixed effects on health care provider time use and a general lack of reliable data to determine financial costs. The authors emphasized the urgent need for research into commercial EHR systems and the adoption of common standards for use in HIT research [46].

A systematic review by Jones et al [47] examined how HIT implementation affected the functionalities (ie, health care quality, safety, and efficiency) described in the meaningful use incentive program in the United States. Of the 236 articles published between 2010 and 2012 included in the review, 170 investigated care quality outcomes, 46 explored the effects of HIT on patient safety, and 62 pertained to efficiency outcomes [47]. Most studies (78%) found that HITs had a positive impact on improving patient safety; however, several papers highlighted that some aspects, such as alert fatigue, could have a negative impact on medication safety and clinical workflows [47].

Although the review by Jones et al [47] review covered HITs instead of solely EHRs, the authors described a similar impression of the HIT literature landscape as that observed in our study: diverse outcome measures, unreliable reporting of findings, and inconsistent study quality, resulting in an unclear understanding of the impact on the quality of care that patients received.

Another noteworthy review was that by Rahurkar et al [48], who examined how HIE use affected health care measures such as costs, use of services, and care quality. The review included 27 articles, with the majority focusing on HIE use in the United States (70%), emergency department settings (52%), and hospitals (26%) [48]. The authors noted that many observational studies reported HIEs as having a positive effect on their outcome measure of interest, with a large proportion of those reporting on quality of care (80%). However, when accounting for study design, especially those using methods with higher internal validity (eg, randomized control trials), the authors found no strong evidence supporting HIE use being causally related to any purported benefits [48].

The review by Reis et al [3] was perhaps the closest attempt to examine how EHR interoperability may affect the 6 facets of care quality. The authors found that eHealth systems with information exchange capabilities could potentially help with the automated detection of health care-acquired infections or patient harm and work efficiency, although benefits such as enhancing documentation accuracy and quality are less clear [3]. The authors identified no studies that investigated the potential cost benefits of interoperability [3]. Although only a limited number of papers met the inclusion criteria and examined an assortment of eHealth systems used, the authors concluded that eHealth systems with interoperability have some positive effects on patient care quality in certain clinical applications (eg, disease, event surveillance) [3].

### Implications for Policy and Further Research

Realizing the well-recognized benefits of interoperable EHRs is often hampered by a multitude of contextual factors. Contributing factors such as the considerable expense associated with introducing new health IT systems and infrastructure, differing procurement policies between health systems, lack of business incentives, and fundamental technical challenges impede the introduction of interventions aimed at addressing interoperability between EHRs [23,49,50].

As demonstrated in the literature, piecemeal interventions aimed at linking EHR systems at a technical level may alleviate bottlenecks in one area of the clinical pathway but only to have the gains undermined elsewhere and culminate in not meaningfully improving the patient’s care during their clinical encounter.

Given that interoperable EHR use by health care workers has already been extensively researched, similar efforts must be devoted to investigating the exact benefits to patient safety or quality of care from the perspective of patients and caregivers. In addition, further research efforts should be devoted to exploring EHR interoperability interventions that do not yield positive results. Understanding these failed interventions and their underlying causes can help better inform our insight into interventions that have shown greater promise.

The adoption of common technical standards and support for regulatory and legislative alignment will prove valuable in further ushering in greater interoperability in the coming years. Likewise, unifying a common set of key performance indicators is also needed to allow for more transparent and comparable metrics for the continued monitoring and evaluation of new policies. Outcomes such as PSEs, medication reconciliation, and time savings may be obvious starting points, but a more consolidated list of universal outcome measures is essential to accurately quantify the effect of these complex interventions.

### Conclusions

Our systematic review found that interoperable EHRs have had a positive impact on certain aspects of patient safety, such as medication reconciliation, reducing PSEs, lowering the risk of data errors, and improving data quality. However, a reliable determination of their true measure of effect with the available assortment of evidence remains difficult. Current evidence underscores the value and importance of continuing to implement greater interoperability in the upgrading of existing EHR systems and the procurement of new ones. In practice, however, relying on present findings to inform exact outcome measure improvements to expect, may prove challenging.

As clinical data increase both in volume and complexity, EHR interoperability will become indispensable for realizing a more streamlined and sustainable workflow for clinicians. Standardizing outcome measures, examining EHR interoperability through the lens of systems complexity, and greater inclusivity of patient perspectives in EHR-related research will be necessary to better evaluate the growing importance of interoperable EHRs in high-income health care settings in the foreseeable future.

## Acknowledgments

This work was supported by the Imperial College National Institute for Health Research Patient Safety Translational Research Centre (reference no PSTRC-2016-004). The funders or sponsors have had no role in the development and drafting of this manuscript.

## Conflicts of Interest

HA is chief scientific officer of Preemptive Health and Medicine at Flagship Pioneering. AD is executive chair of Preemptive Health and Medicine at Flagship Pioneering.

Multimedia Appendix 1
Main outcomes table.

[DOCX File , 54 KB - jmir_v24i9e38144_app1.docx ]

---

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 69 KB - jmir_v24i9e38144_app2.pdf ]

---

## References

1. Hersh WR, Totten AM, Eden KB, Devine B, Gorman P, Kassakian SZ, et al. Outcomes from health information exchange: systematic review and future research needs. JMIR Med Inform 2015 Dec 15;3(4):e39 [FREE Full text] [doi: 10.2196/medinform.5215] [Medline: 26678413]
2. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2018 Mar 06;25(S 01):S48-S61. [doi: 10.15265/iys-2016-s006]
3. Reis ZS, Maia TA, Marcolino MS, Becerra-Posada F, Novillo-Ortiz D, Ribeiro AL. Is there evidence of cost benefits of electronic medical records, standards, or interoperability in hospital information systems? Overview of systematic reviews. JMIR Med Inform 2017 Aug 29;5(3):e26 [FREE Full text] [doi: 10.2196/medinform.7400] [Medline: 28851681]
4. Interoperability in Healthcare. HIMSS. URL: https://www.himss.org/resources/interoperability-healthcare [accessed 2022-02-04]
5. Begoyan A. An overview of interoperaiblity standards for electronic health records. Integrated Design and Process Technology. 2007. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.4421&rep=rep1&type=pdf [accessed 2021-12-07]
6. Studeny J, Coustasse A. Personal health records: is rapid adoption hindering interoperability? Perspect Health Inf Manag 2014;11:1e [FREE Full text] [Medline: 25214822]
7. Thompson MP, Graetz I. Hospital adoption of interoperability functions. Healthc (Amst) 2019 Sep;7(3):100347. [doi: 10.1016/j.hjdsi.2018.12.001] [Medline: 30595558]
8. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. P T 2017 Sep;42(9):572-575 [FREE Full text] [Medline: 28890644]
9. Glickman M, Orlova A. Building interoperability standards and ensuring patient safety. J AHIMA 2015;86(11):48-51. [Medline: 26910934]
10. Goals and benefits of data interoperability. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/datainteroperability/goals-and-benefit.html [accessed 2022-02-04]
11. Warren LR, Clarke J, Arora S, Darzi A. Improving data sharing between acute hospitals in England: an overview of health record system distribution and retrospective observational analysis of inter-hospital transitions of care. BMJ Open 2019 Dec 05;9(12):e031637 [FREE Full text] [doi: 10.1136/bmjopen-2019-031637] [Medline: 31806611]
12. Pendergrass JC, Chandrasekaran R. Key factors affecting ambulatory care providers' electronic exchange of health information with affiliated and unaffiliated partners: web-based survey study. JMIR Med Inform 2019 Nov 07;7(4):e12000 [FREE Full text] [doi: 10.2196/12000] [Medline: 31697241]
13. Everson J, Adler-Milstein J. Gaps in health information exchange between hospitals that treat many shared patients. J Am Med Inform Assoc 2018 Sep 01;25(9):1114-1121 [FREE Full text] [doi: 10.1093/jamia/ocy089] [Medline: 30010887]
14. Neves AL, Freise L, Laranjo L, Carter AW, Darzi A, Mayer E. Impact of providing patients access to electronic health records on quality and safety of care: a systematic review and meta-analysis. BMJ Qual Saf 2020 Dec 12;29(12):1019-1032 [FREE Full text] [doi: 10.1136/bmjqs-2019-010581] [Medline: 32532814]
15. Six domains of health care quality. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/talkingquality/measures/six-domains.html [accessed 2022-02-03]
16. Carayon P, Wood K. Patient safety - the role of human factors and systems engineering. Stud Health Technol Inform 2010;153:23-46 [FREE Full text] [Medline: 20543237]
17. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm A New Health System for the 21st Century. Washington, D.C., United States: National Academies Press; 2001.
18. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 2007 Mar 28;297(8):831-841. [doi: 10.1001/jama.297.8.831] [Medline: 17327525]
19. Clarke JM, Warren LR, Arora S, Barahona M, Darzi AW. Guiding interoperable electronic health records through patient-sharing networks. NPJ Digit Med 2018 Dec 10;1(1):65 [FREE Full text] [doi: 10.1038/s41746-018-0072-y] [Medline: 31304342]
20. Coleman E, Berenson R. Lost in transition: challenges and opportunities for improving the quality of transitional care. Ann Intern Med 2004 Oct 05;141(7):533-536. [doi: 10.7326/0003-4819-141-7-200410050-00009] [Medline: 15466770]
21. Ngugi P, Were M, Babic A. Facilitators and barriers of electronic medical records systems implementation in low resource settings: a holistic view. Stud Health Technol Inform 2018;251:187-190. [Medline: 29968634]

22. Lyles C, Schillinger D, Sarkar U. Connecting the dots: health information technology expansion and health disparities. PLoS Med 2015 Jul 14;12(7):e1001852 [FREE Full text] [doi: 10.1371/journal.pmed.1001852] [Medline: 26172977]
23. Wilson K, Khansa L. Migrating to electronic health record systems: a comparative study between the United States and the United Kingdom. Health Policy 2018 Nov;122(11):1232-1239. [doi: 10.1016/j.healthpol.2018.08.013] [Medline: 30193980]
24. Clarke A, Watt I, Sheard L, Wright J, Adamson J. Implementing electronic records in NHS secondary care organizations in England: policy and progress since 1998. Br Med Bull 2017 Jan 01;121(1):95-106. [doi: 10.1093/bmb/ldw055] [Medline: 28043952]
25. Li E, Clarke J, Neves AL, Ashrafian H, Darzi A. Electronic health records, interoperability and patient safety in health systems of high-income countries: a systematic review protocol. BMJ Open 2021 Jul 14;11(7):e044941 [FREE Full text] [doi: 10.1136/bmjopen-2020-044941] [Medline: 34261679]
26. World Bank country and lending groups. The World Bank. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519 [accessed 2021-05-26]
27. Khokhar T, Serajuddin U. Should we continue to use the term “developing world”? World Bank Blogs. 2015 Nov 16. URL: https://blogs.worldbank.org/opendata/should-we-continue-use-term-developing-world [accessed 2021-05-26]
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. 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]
29. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/bm.2012.031]
30. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
31. Elysee G, Herrin J, Horwitz L. An observational study of the relationship between meaningful use-based electronic health information exchange, interoperability, and medication reconciliation capabilities. Medicine (Baltimore) 2017 Oct;96(41):e8274 [FREE Full text] [doi: 10.1097/MD.0000000000008274] [Medline: 29019898]
32. Motulsky A, Couture I, Weir D, Tamblyn R. Incorporating pharmacy dispensing records into medical records: usability challenges. In: Studies in Health Technology and Informatics. Amsterdam, Netherlands: IOS Press; 2021.
33. Lee Y, Chien T, Hsu S, Chang P. Developing and applying a cross-disciplinary team handover information system. In: Studies in Health Technology and Informatics. Amsterdam, Netherlands: IOS Press; 2013.
34. Reed M, Huang J, Brand R, Graetz I, Jaffe MG, Ballard D, et al. Inpatient-outpatient shared electronic health records: telemedicine and laboratory follow-up after hospital discharge. Am J Manag Care 2020 Oct 01;26(10):e327-e332 [FREE Full text] [doi: 10.37765/ajmc.2020.88506] [Medline: 33094945]
35. Wong SP, Jacobson HN, Massengill J, White HK, Yanamadala M. Safe interorganizational health information exchange during the COVID-19 pandemic. J Am Med Dir Assoc 2020 Dec;21(12):1808-1810 [FREE Full text] [doi: 10.1016/j.jamda.2020.10.022] [Medline: 33162358]
36. D'Amore J, Bouhaddou O, Mitchell S, Li C, Leftwich R, Turner T, et al. Interoperability progress and remaining data quality barriers of certified health information technologies. AMIA Annu Symp Proc 2018;2018:358-367 [FREE Full text] [Medline: 30815075]
37. Akbarov A, Kontopantelis E, Sperrin M, Stocks SJ, Williams R, Rodgers S, et al. Primary care medication safety surveillance with integrated primary and secondary care electronic health records: a cross-sectional study. Drug Saf 2015 Jul 23;38(7):671-682 [FREE Full text] [doi: 10.1007/s40264-015-0304-x] [Medline: 26100143]
38. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic health record usability issues and potential contribution to patient harm. JAMA 2018 Mar 27;319(12):1276-1278 [FREE Full text] [doi: 10.1001/jama.2018.1171] [Medline: 29584833]
39. Adams K, Howe J, Fong A, Puthumana J, Kellogg K, Gaunt M, et al. An analysis of patient safety incident reports associated with electronic health record interoperability. Appl Clin Inform 2017 Dec 21;08(02):593-602. [doi: 10.4338/aci-2017-01-ra-0014]
40. Biltoft J, Finneman L. Clinical and financial effects of smart pump-electronic medical record interoperability at a hospital in a regional health system. Am J Health Syst Pharm 2018 Jul 15;75(14):1064-1068. [doi: 10.2146/ajhp161058] [Medline: 29987060]
41. Koldby S, Schou JI. Clinical simulation and workflow by use of two clinical information systems, the electronic health record and digital dictation. In: Studies in Health Technology and Informatics. Amsterdam, Netherlands: IOS Press; 2013.
42. Munck LK, Hansen KR, Mølbak AG, Balle H, Kongsgren S. The use of shared medication record as part of medication reconciliation at hospital admission is feasible. Dan Med J 2014 May;61(5):A4817 [FREE Full text] [Medline: 24814735]
43. Connecting health and care for the nation a shared nationwide interoperability roadmap. The Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf [accessed 2022-01-07]
44. D'Amore JD, Mandel JC, Kreda DA, Swain A, Koromia GA, Sundareswaran S, et al. Are Meaningful Use Stage 2 certified EHRs ready for interoperability? Findings from the SMART C-CDA Collaborative. J Am Med Inform Assoc 2014;21(6):1060-1068 [FREE Full text] [doi: 10.1136/amiajnl-2014-002883] [Medline: 24970839]

45. Dobrow MJ, Bytautas JP, Tharmalingam S, Hagens S. Interoperable electronic health records and health information exchanges: systematic review. JMIR Med Inform 2019 Jun 06;7(2):e12607 [FREE Full text] [doi: 10.2196/12607] [Medline: 31172961]
46. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
47. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 07;160(1):48-54. [doi: 10.7326/m13-1531]
48. Rahurkar S, Vest JR, Menachemi N. Despite the spread of health information exchange, there is little evidence of its impact on cost, use, and quality of care. Health Aff (Millwood) 2015 Mar;34(3):477-483. [doi: 10.1377/hlthaff.2014.0729] [Medline: 25732499]
49. Rozenblum R, Jang Y, Zimlichman E, Salzberg C, Tamblyn M, Buckeridge D, et al. A qualitative study of Canada's experience with the implementation of electronic health information technology. Can Med Assoc J 2011 Mar 22;183(5):E281-E288 [FREE Full text] [doi: 10.1503/cmaj.100856] [Medline: 21343262]
50. Bowden T, Coiera E. Comparing New Zealand's 'Middle Out' health information technology strategy with other OECD nations. Int J Med Inform 2013 May;82(5):e87-e95. [doi: 10.1016/j.ijmedinf.2012.12.002] [Medline: 23287413]

## Abbreviations

**EHR:** electronic health record
**HIC:** high-income country
**HIE:** health information exchange
**HIT:** health information technology
**IoM:** Institute of Medicine
**PICO:** population, intervention, control, and outcomes
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PSE:** patient safety event

*Edited by A Mavragani; submitted 21.03.22; peer-reviewed by Z Reis, C Sullivan, Z Mahmoudvand, K Matthias; comments to author 15.06.22; revised version received 18.07.22; accepted 23.08.22; published 15.09.22.*

*Please cite as:*
*Li E, Clarke J, Ashrafian H, Darzi A, Neves AL*
*The Impact of Electronic Health Record Interoperability on Safety and Quality of Care in High-Income Countries: Systematic Review*
*J Med Internet Res 2022;24(9):e38144*
*URL: https://www.jmir.org/2022/9/e38144*
*doi:10.2196/38144*
*PMID:36107486*

Review

# Social Media Use and Health and Well-being of Lesbian, Gay, Bisexual, Transgender, and Queer Youth: Systematic Review

Matthew N Berger[1,2], BSc, GCEdSt, MPhil, MPH, MScMed; Melody Taba[1], BSc (Hons), MPH; Jennifer L Marino[1,3,4,5,6,7], PhD; Megan S C Lim[7,8,9], PhD; S Rachel Skinner[1,10], MBBS, PhD, FRACP

[1]Specialty of Child and Adolescent Health, The Children’s Hospital at Westmead Clinical School, Faculty of Medicine and Health, The University of Sydney, Westmead, Australia

[2]Centre for Population Health, Western Sydney Public Health Unit, North Parramatta, Australia

[3]Department of Obstetrics and Gynaecology, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, Australia

[4]Royal Women's Hospital, Parkville, Australia

[5]Centre for Adolescent Health, Murdoch Children’s Research Institute, Parkville, Australia

[6]Department of Paediatrics, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, Australia

[7]Centre for Epidemiology and Biostatistics, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, Australia

[8]Burnet Institute, Melbourne, Australia

[9]School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[10]Kids Research, Children’s Hospital Westmead, Sydney Children’s Hospitals Network, Westmead, Australia

**Corresponding Author:**
Matthew N Berger, BSc, GCEdSt, MPhil, MPH, MScMed
Specialty of Child and Adolescent Health, The Children’s Hospital at Westmead Clinical School
Faculty of Medicine and Health
The University of Sydney
Corner Hawkesbury Road and, Hainsworth St
Westmead, 2145
Australia
Phone: 61 298403603
Email: matthew.berger@health.nsw.gov.au

## Abstract

**Background:** Lesbian, gay, bisexual, transgender, and queer (LGBTQ) individuals are at higher risk of poor mental health and well-being. Social media platforms can provide LGBTQ youths with a space that counters heteronormative environments and potentially supports mental health and well-being. Mental health includes an individual’s state of psychological and emotional well-being and not merely the absence of mental disorders.

**Objective:** We sought to identify how LGBTQ youths and adolescents use social media for connection with other LGBTQ peers and groups, identity development, and social support and how these affect mental health and well-being.

**Methods:** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) procedures were used to guide this review. Searches were conducted in ACM Digital Library, CINAHL, Ovid Embase, Ovid MEDLINE, and Web of Science in March 2021. This review focused on LGBTQ youths aged 10 to 24 years. Included peer-reviewed studies must comprise social media; explore peer connection, identity development, or social support; and be published from 2012 onward. In total, 2 researchers extracted data and performed quality assessments independently using the Newcastle-Ottawa Scale for quantitative articles and the Critical Appraisal Skills Programme for qualitative articles. Qualitative synthesis was performed on articles that satisfied the eligibility criteria.

**Results:** A total of 26 studies (n=15, 58% qualitative; n=8, 31% quantitative; n=3, 12% mixed methods) met the inclusion criteria. Of the 8 quantitative studies, 6 (75%) were cross-sectional, and 2 (25%) were cohort studies. All studies ranged from moderate to high quality. Social media was a popular tool used by LGBTQ youths to connect with LGBTQ communities. In qualitative data, we found that LGBTQ youths negotiated and explored identity and obtained support from peers on social media. Instagram, Tumblr, and Twitter were commonly used to access LGBTQ content owing to ease of anonymity. Identity management was the most studied social media affordance, important to LGBTQ youths for strategic disclosure. Key strategies for managing identities included being anonymous, censoring locations or content, restricting audiences, and using multiple accounts. Quantitative studies (3/8, 38%) showed that social media was associated with reduced mental health concerns and increased well-being among LGBTQ youths. Mental health concerns arising from social media use were attributed to discrimination, victimization, and policies that did not accommodate changed identities.

**Conclusions:** We found that social media may support the mental health and well-being of LGBTQ youths through peer connection, identity management, and social support, but findings were limited by weaknesses in the evidence. More robust and longitudinal studies are needed to determine the relationship between social media use and LGBTQ mental health, particularly among adolescents. The findings may inform interventions to promote social media health literacy and the mental health and well-being of this vulnerable group.

**Trial Registration:** PROSPERO CRD42020222535; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=222535

***(J Med Internet Res 2022;24(9):e38449)*** doi:10.2196/38449

**KEYWORDS**

lesbian, gay, bisexual, transgender, and queer; LGBTQ; adolescence; youth; well-being; mental health; social media; identity; support; mobile phone

# *Introduction*

## Background

Recent years have seen social media become a part of our daily lives, especially for adolescents and young adults [1]. Facebook, Instagram, Snapchat, TikTok, and YouTube are among the most popular platforms used by adolescents [2,3]. Social media use can be defined as web-based behaviors using platforms to like, comment, message, or monitor other users [4]. Social media can be used to overcome barriers of distance and expand or consolidate web-based communities [5]. Several benefits to well-being have been associated with social media, including strengthened peer relationships, involvement in specific social networks, and facilitation of identity expression [6,7]. Social media platforms are constantly evolving and encourage a plethora of activities ranging from communicating with family and friends to sharing content and knowledge [8]. Motivations for social media use include entertainment, relationships, information, and identity development and management [9]. Lesbian, gay, bisexual, transgender, and queer (or questioning; LGBTQ) people are heavier users of social media and are more likely to have multiple accounts compared with their non-LGBTQ counterparts [10].

LGBTQ people experience higher rates of mental health concerns and behaviors, including suicidal ideation, self-harm, anxiety, depression, and posttraumatic stress disorder [11]. LGBTQ populations are also at a higher risk of experiencing violence, discrimination, and adversity [12-14]. LGBTQ youths in particular have a higher prevalence of victimization than non-LGBTQ youths because of increased exposure to prejudice and violence at school [13]. Unsupportive family and peers contribute significantly to an increased risk of mental health disorders and substance use [15-17]. However, in some situations, disclosure of sexual or gender identity, or “coming out,” is associated with reduced mental health issues [15-17]. To counter the negative consequences of coming out, some LGBTQ individuals use selective disclosure strategies, particularly because of concerns about losing friends or family [16]. Family, friend, and society acceptance are associated with better mental health, well-being, and self-esteem in LGBTQ individuals [18,19]. Other support networks such as involvement in LGBTQ sporting clubs can also improve mental health and well-being among LGBTQ people [20,21]. In addition, LGBTQ people tend to rely on other LGBTQ individuals for support [10,20,21], and not connecting with LGBTQ support networks is associated with poorer mental health outcomes [20].

Many LGBTQ individuals live in environments where sexuality and gender diversity are not accepted [11]. At least 69 countries criminalize same-sex relationships, and 9 countries criminalize gender nonconformity [22]. These environments make LGBTQ identity development difficult and, in public, individuals are forced to conform to heteronormativity to avoid persecution. Even societies that are more accepting of LGBTQ people maintain mainstream heteronormative environments [23]. For example, school sex education focuses on heterosexual people and is rarely inclusive of other sexualities [24,25]. Social media can act as a safe environment to access information about identity, express identity, or provide support among LGBTQ people, thus supporting mental health and well-being [26-34]. Although individual studies have shown benefits, there has not been a review of studies to synthesize the range of benefits. It is important to understand how social media is used by LGBTQ youths to explore their identities and connect with like-minded people and how this affects mental health. Thus, this highlights the important role that social media plays regarding LGBTQ youth for policy makers, educators, and clinicians working in this area. This population is an important focus for research because of the increased risk of compromised mental health and well-being [11].

## Aims

In this systematic review, we sought to examine studies exploring the relationship between social media use and mental health and well-being among LGBTQ youths. Specifically, we aimed to identify how LGBTQ youths and adolescents use social media for (1) connection with other LGBTQ peers and groups, (2) identity development, and (3) social support and how these affect mental health and well-being. We also sought to identify any impact of social media on the mental health of LGBTQ youths. The World Health Organization classifies young people as those aged between 10 and 24 years [35]. For the purpose of this review, “youth” includes adolescent and youth ages.

# Methods

## Registration and Search Strategy

This review was registered with PROSPERO before data synthesis (CRD42020222535; Multimedia Appendix 1). Electronic databases were searched for literature, including CINAHL (1939; March 2021), Ovid Embase (1947; March 2021), Ovid MEDLINE (1946; March 2021), Web of Science (1900; March 2021), and ACM Digital Library (1985; March 2021). Additional studies were found through Google Scholar and PubMed and added to the screening process. A hand search of the reference lists of the included papers was also conducted to identify any studies missed in the search terms. Individual database searches are listed in Multimedia Appendix 2. These searches were conducted using a search strategy with the following keywords: *LGB* or GLB* or Sexual and Gender Minorities or gay or lesbian or queer or transgender or sexually and gender diverse or gender and sexually diverse or homosexual* or bisexual* or sexual orientation AND identit* or support* or help* or friend* or relationship* or partner* or mental health or depression or anxiety or mood disorder or posttraumatic stress disorder or PTSD or suicid* or self-harm or wellbeing AND social media* or social networking site* or Facebook or Instagram or Tumblr or Twitter* or YouTube or LinkedIn or WeChat or Snapchat or TikTok AND adolescen* or young adult* or teen* or youth**.

## Inclusion and Exclusion Criteria

To be included, studies needed to (1) have a sample consisting of at least 50% individuals aged 10 to 24 years to ensure that the study focused on youths; (2) be specific to LGBTQ populations or present LGBTQ findings separately from any non-LGBTQ sample; (3) include social media use as a predictor; (4) explore connecting with peers, identity development, or social support; (5) be published from 2012 onward to capture recent forms of social media platforms and use (including smartphone apps); and (6) be available in full text and in English. All study designs were eligible, including quantitative, qualitative, and mixed methods research. Only peer-reviewed articles of original research were eligible; case studies, narratives, conference presentations, and other nonempirical works were not included. Papers were first screened by title and abstract (961/1234, 77.88%) and again by full text (101/961, 10.5%) by MNB and MT. A total of 26 papers were appraised by MNB and MT. Disagreements during title and abstract screening and full-text assessments were discussed between MNB and MT, and any disagreements were resolved through team discussion.

## Quality Assessment

All the included studies were subject to quality appraisal to assess the research design, ethics compliance, and risk of bias. The Newcastle-Ottawa Quality Assessment Scale (NOS) was used to assess the quality of quantitative studies [36], including an adapted version for cross-sectional studies [37]. The NOS assesses studies based on 3 domains: selection, comparability, and outcome [36]. For qualitative studies, the Critical Appraisal Skills Programme (CASP) was used to assess quality [38]. Both the NOS and CASP were applied to mixed methods studies. Mixed methods studies are discussed in the relevant qualitative or quantitative sections.

## Data Synthesis

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) procedures were used to guide the review (Multimedia Appendix 3) [39]. A quality assessment table displaying CASP or NOS scores and a summary table of the studies, including study characteristics, were produced. As quantitative studies used different measures for outcomes and took different statistical approaches, a meta-analysis was not possible. For qualitative data, a predefined schema was developed to assist with data collection based on the review's aims before conducting the search according to preliminary literature searches. Included qualitative data were thematically synthesized according to the 3 stages of Thomas and Harden [40], namely, coding, developing and refining themes, and generating analytical themes. The findings were divided into 3 themes and, within each theme, into qualitative and quantitative findings. The three themes as developed by the schema were (1) connecting with other LGBTQ youths on social media, (2) LGBTQ identity development using social media, and (3) social support on social media. For this review, "queer" represents gender or sexualities otherwise not classified within lesbian, gay, bisexual, and transgender. Identity development refers to the exploration of a diverse sexuality or gender, which includes discovery, awareness, appraisal, communication, and how individuals manage their identity [41]. Social support via social media among LGBTQ youths refers to receiving assistance and feeling cared for [26]. These are important factors with the potential to promote mental health and well-being when individuals can explore and connect in safe spaces [42].

# Results

## Overview

This search resulted in a total of 961 papers retrieved from the specified databases, with 273 (28.4%) duplicates removed (Figure 1). Title and abstract screening excluded 89.5% (860/961) of the papers, leaving 101 papers for full-text screening, of which 26 (25.7%) met the aims and criteria of this review. Of the 26 included papers, 15 (58%) were qualitative studies, 8 (31%) were quantitative studies, and 3 (12%) were mixed methods studies. The included studies were mostly conducted in the United States (17/26, 65%), whereas others were conducted in Australia (2/26, 8%), Canada (4/26, 15%), China (1/26, 4%), Ukraine (1/26, 4%), and the United Kingdom (3/26, 12%). The ages of the study participants ranged from 13 to 34 years, with a total of 14,112 participants across the 26 studies. One study appeared to meet the inclusion criteria, but we were unable to confirm the age descriptions of their sample through the full text or contacting the authors [23]. Summaries of quality assessments are provided in Table 1, and summaries of included qualitative and quantitative studies are provided in Table 2 and Multimedia Appendix 4.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the selection process. LGBTQ: lesbian, gay, bisexual, transgender, and queer.

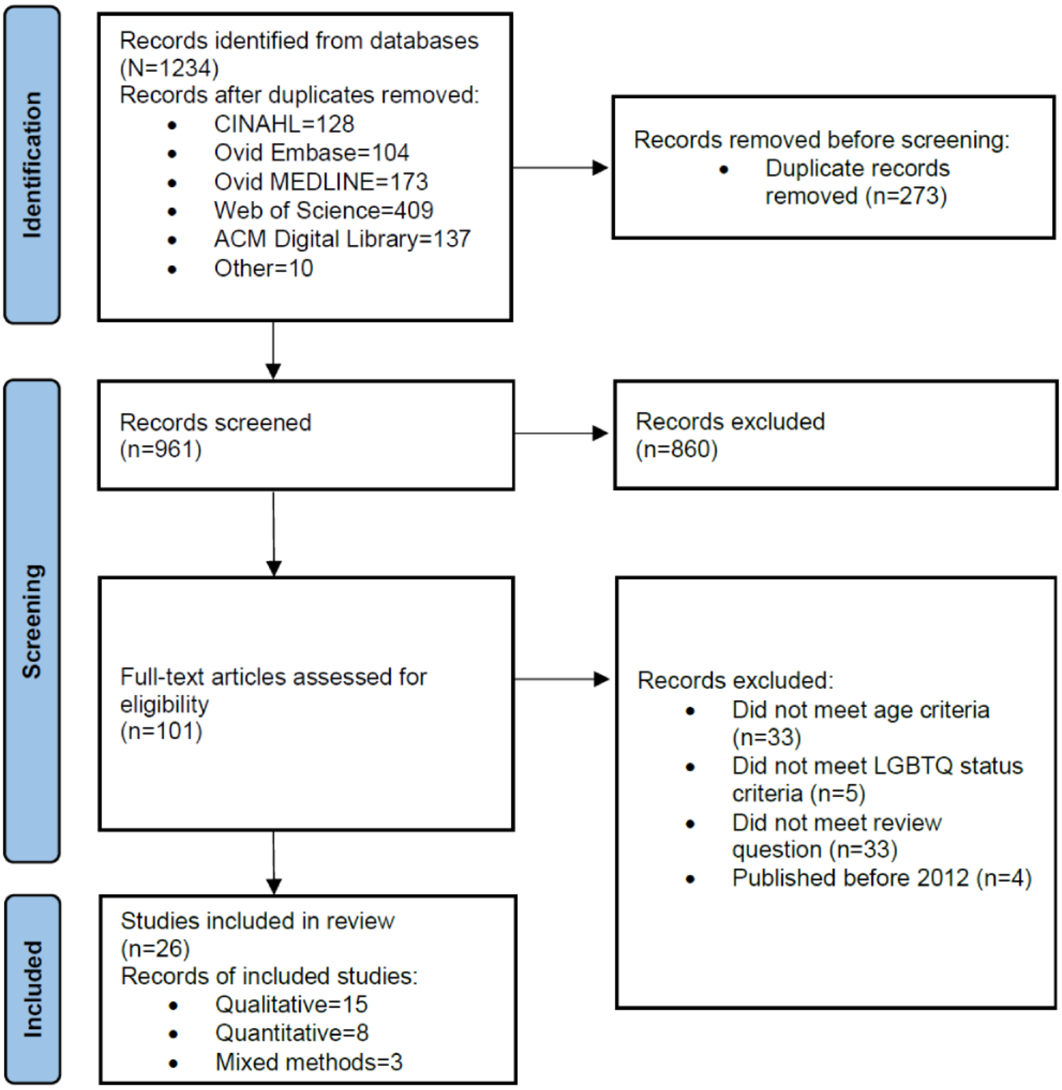


**Table 1.** Quality assessment summaries and limitations of the included studies (N=26).

| Study, year | CASP[a] score | NOS[b] score | Comments and limitations |
|---|---|---|---|
| Bates et al [27], 2020 | 8/10 criteria | N/A[c] | • Generalizability:<br>• Participants predominately White and openly LGBTQ[d]<br>• All recruited from 1 university |
| Bond and Figueroa-Caballero [43], 2016 | N/A | 8/10 stars | • Cross-sectional study<br>• Generalizability:<br>• Data collected from gay-straight alliances |
| Byron et al [31], 2019[e] | 8/10 criteria | 5/10 stars | • Generalizability:<br>• Race and ethnicity not well described<br>• Internal validity:<br>• Possible risk of interviewer bias not described |
| Ceglarek and Ward [44], 2016 | N/A | 9/10 stars | • Cross-sectional study<br>• Internal validity:<br>• Data were self-reported, which may be prone to social desirability or recall bias<br>• Inadequately validated measures |
| Chong et al [45], 2015 | N/A | 7/10 stars | • Cross-sectional study<br>• Generalizability:<br>• Small Hong Kong–based LGB[f] population |
| Craig and McInroy [32], 2014 | 9/10 criteria | N/A | • Generalizability:<br>• Most participants were from progressive, well-educated, and affluent backgrounds |
| Craig et al [46], 2021 | N/A | 6/10 stars | • Cross-sectional study<br>• Internal validity:<br>• Inadequately validated measures<br>• Sample characteristics not described (covariates) |
| Duguay [47], 2016 | 8/10 criteria | N/A | • Generalizability:<br>• Small gender-diverse population within the sample<br>• All participants were university students |
| Fox and Ralston [48], 2016 | 9/10 criteria | N/A | • Generalizability:<br>• Participants predominately White<br>• From 1 city in the United States, with most being college students<br>• Internal validity:<br>• Possible risk of interviewer bias not described |
| Hanckel et al [49], 2019 | 7/10 criteria | N/A | • Generalizability:<br>• Suboptimal description of race and ethnicity but indicating a lack of diversity<br>• Internal validity:<br>• Selection and recruitment not described<br>• Analytical method unclear |
| Harper et al [33], 2016 | 10/10 criteria | N/A | • Generalizability:<br>• Data were collected from 2004 to 2006 and, thus, may not represent current use and past perceptions of LGBTQ identities<br>• Only recruited from 2 metropolitan US cities<br>• Limited ethnic backgrounds because of the parent study aims |

| Study, year | CASP[a] score | NOS[b] score | Comments and limitations |
|---|---|---|---|
| Herrera [50], 2018 | 8/10 criteria | N/A | • Generalizability:<br>• No participant characteristics described<br>• Limited to lesbian or queer-identifying women<br>• Limited to Instagram use<br>• Internal validity:<br>• Concern of selection and interviewer bias as the investigators invited participants to interviews by commenting on Instagram posts |
| Hillier et al [28], 2012 | 9/10 criteria | N/A | • Generalizability:<br>• Participants predominately White |
| Lucero [29], 2017[e] | 8/10 criteria | 5/10 stars | • Generalizability:<br>• Small sample of Ukrainian youths that may not be representative of the Ukrainian population<br>• Investigators reported their results as LGBTQ although there were no transgender participants<br>• Internal validity:<br>• Significantly small sample size for quantitative analysis<br>• Inadequately validated measures |
| McConnell et al [51], 2018[e] | 10/10 criteria | 6/10 stars | • Generalizability:<br>• From a single metropolitan US city<br>• The disclosed LGBTQ identity cohort were likely overrepresented because of the significantly higher sample size compared with nondisclosed cohorts<br>• Participants predominately African American |
| McConnell et al [52], 2017 | N/A | 6/10 stars | • Prospective cohort study<br>• Generalizability:<br>• Participants predominately African American<br>• Internal validity:<br>• Loss to follow-up not described |
| McInroy et al [30], 2019 | N/A | 7/10 stars | • Cross-sectional study<br>• Internal validity:<br>• Possible selection bias as the study aimed to compare on the web and offline; however, it recruited primarily on the web |
| McInroy and Craig [53], 2015 | 8/10 criteria | N/A | • Generalizability:<br>• From a single metropolitan Canadian city<br>• Small transgender subpopulation<br>• Internal validity:<br>• Most had high motivation or knowledge of media and may be associated with volunteer bias |
| Paceley et al [54], 2020 | 9/10 criteria | N/A | • Generalizability:<br>• From small towns or rural areas in 1 US state<br>• Although they purposefully recruited diverse participants, there was limited intersectional analysis<br>• Underrepresentative of transgender population |
| Pellicane et al [55], 2020 | N/A | 5/10 stars | • Prospective cohort study<br>• Generalizability:<br>• Undergraduate psychology students from 1 university<br>• Participants predominately female<br>• Internal validity:<br>• Strong risk of volunteer bias because of selection<br>• Loss to follow-up not described |

| Study, year | CASP[a] score | NOS[b] score | Comments and limitations |
|---|---|---|---|
| Rubin and McClelland [34], 2015 | 9/10 criteria | N/A | • Generalizability:<br>  • Most from a single metropolitan US city<br>• Internal validity:<br>  • Small sample size |
| Selkie et al [56], 2020 | 9/10 criteria | N/A | • Generalizability:<br>  • From 1 gender services clinic in the Midwestern United States<br>  • All had supportive parents because of recruitment from the clinic<br>• Internal validity:<br>  • Participants' locality not collected (eg, rural or metropolitan) |
| Singh [57], 2013 | 9/10 criteria | N/A | • Internal validity:<br>  • Risk of bias from telephone interviews |
| Taylor et al [58], 2014 | 10/10 criteria | N/A | • Generalizability:<br>  • Underrepresentative of transgender population |
| Twist et al [59], 2017 | N/A | 4/10 stars | • Cross-sectional study<br>• Generalizability:<br>  • Undergraduate students minoring in Family Studies from 1 university<br>• Internal validity:<br>  • Strong risk of volunteer bias because of selection<br>  • Small sample size<br>  • Data were self-reported, which may be prone to social desirability or recall bias |
| Varjas et al [60], 2013 | 9/10 criteria | N/A | • Generalizability:<br>  • Owing to age, parental permission was required, and the study likely included only those whose parents knew and were supportive |

[a] CASP: Critical Appraisal Skills Programme.

[b] NOS: Newcastle-Ottawa Quality Assessment Scale.

[c] N/A: not applicable.

[d] LGBTQ: lesbian, gay, bisexual, transgender, and queer.

[e] Mixed methods studies.

[f] LGB: lesbian, gay, and bisexual.

**Table 2.** Summary of the included quantitative studies (N=11).

| Study, year, and country | Purpose | Age (years) | Sample size, N | LGBTQ[a] sample | Method | Findings | Summary |
|---|---|---|---|---|---|---|---|
| Bond and Figueroa-Caballero [43], 2016, United States | Understand the relationships among technology, sexual identity, and well-being based on age, gender, geographic location, race, and religion | 13 to 19 (mean 16.5, SD 1.3) | 570 | Gay (45%), bisexual (27%), and lesbian (24%) | Recruitment: from gay-straight alliances and web-based message boards; data collection: questionnaires and surveys; measures: Rosenberg Self-Esteem Scale, Multiple Affect Adjective Check List, Multidimensional Scale of Perceived Social Support, and Measure of Sexual Identity Exploration and Commitment | Using regression analyses, the study found that LGB[b] youths spend more time on social media compared with non-LGB youths, with time spent significantly on sexual identity ($\beta=.14$; $P<.05$) and well-being ($\beta=.11$; $P<.05$). Time spent on social media was associated with sexual identity ($\beta=.08$; SE 0.02; $P<.001$) but not directly with well-being ($\beta=.04$; SE 0.03; $P=.21$). Well-being was significantly associated with sexual identity commitment ($\beta=.47$; SE 0.08; $P<.001$). | Social media demonstrated a connection with sexual identity development associated with well-being. LGBTQ youths used social media to understand sexuality and give social support, which may not be as significant offline. |
| Byron et al [31], 2019, Australia | How Tumblr is used among LGBTQ youths to connect with peers and develop identity and well-being | 16 to 34 (mean 24.6) | 1304 | Homosexual (33.9%); bisexual (24.7%); queer (18%); and pansexual, agender, panromantic, and demisexual (19.8%) | Recruitment: via social media advertisements and flyers to LGBTQ organizations; data collection: semistructured interviews and questionnaires and surveys; measures: 2 nominal questions | Tumblr was the platform that participants most often left (11.7%; excluding Myspace and Tinder). Tumblr was abandoned for several reasons: 34% found it too time-consuming, 30% felt it became a negative space, and 15% found it to have negative health impacts. | Negative experiences were common, with participants describing Tumblr as becoming toxic, although it was useful. |
| Ceglarek and Ward [44], 2016, United States (Michigan) | Understand LGB use of social media for identity exploration and expression and connecting with LGB communities | 18 to 24 (mean 20.23, SD 1.68; LGBTQ participants) | 570 | Heterosexual (n=446), homosexual (n=68), not sure (n=4), and other (n=21) | Recruitment: from LGBTQ support organizations; data collection: questionnaires and surveys; measures: Lesbian, Gay, and Bisexual Identity Scale as well as the Short Scale for Measuring Loneliness in Large Surveys and the Brief Symptom Inventory | Among LGB youths, higher social support on social media was associated with lower levels of loneliness ($\beta=-0.27$; $P\leq.01$) and paranoia ($\beta=-0.21$; $P\leq.05$) using $\beta$ coefficients. There were no significant differences for anxiety ($\beta=-0.03$; $P>.05$), depression ($\beta=-0.10$; $P>.05$), hostility ($\beta=-0.04$; $P>.05$), and sensitivity ($\beta=-0.04$; $P>.05$). Learning about sexuality via social media reduced anxiety ($\beta=-0.35$; $P>.05$), hostility ($\beta=-0.32$; $P\leq.05$), and paranoia ($\beta=-0.43$; $P\leq.01$). | Social media has potential to allow LGBTQ youths to develop identity and, thus, have improved mental health. When seeking identity expressions and social support, the web may provide avenues with reduced stigmatization compared with offline. |

| Study, year, and country | Purpose | Age (years) | Sample size, N | LGBTQ[a] sample | Method | Findings | Summary |
|---|---|---|---|---|---|---|---|
| Chong et al [45], 2015, China (Hong Kong) | Understand LGB social media use for identity, community monitoring, and support and sense of belonging | Mean 23.3 (SD 6.33) | 233 | Lesbian (n=86), gay (n=107), and bisexual (n=40) | Recruitment: flyers distributed to LGBTQ organizations and social media; data collection: questionnaires and surveys; measures: Inclusion of Community in Self Scale, Mental Health Inventory, Life Satisfaction Scale, and Satisfaction with Life Scale | Using structural equation modeling, sense of belonging among LGB youths was associated with social media use for LGB group membership (β=.22; *P*<.05). LGB group connection via social media was indirectly associated with improved mental well-being through reduced stigma (β=.27; *P*<.05). Social media use to enhance LGB connection and reduce stigma affected mental well-being (β=.06 and 0.09; *P*<.05). | Social media is a vital resource for LGB youths to express sexual or gender identity and social support. Mental health can be improved with positive social media capital. |
| Craig et al [46], 2021, Canada and United States | Explore benefits of social media among LGBTQ youths and develop the Social Media Benefits Scale | 14 to 29 (mean 18.21, SD 3.6) | 6178 | Pansexual (n=1782), bisexual (n=1602), queer (n=1305), gay (n=970), lesbian (n=968), asexual (n=691), not sure (n=398), cisgender (n=3950), gender nonconforming (n=2168), and transgender (n=909) | Recruitment: flyers displayed on the web on social media and sent to LGBTQ organizations; data collection: questionnaires and surveys; measures: Social Media Benefits Scale | Of those who chose Facebook as their favorite platform, 11% reported that it helped them feel loved. Adolescents (aged 14-18 years) were the most likely group, and those aged 19 to 24 years were the second most likely group, to use social media for emotional support and development (*F*=75.88; *P*<.001). | Younger youths were more likely to use social media for its benefits, such as social support, connectivity, and information. Youths would commonly connect with LGBTQ individuals or groups and celebrities. Other benefits included improved emotional support and development. |
| Lucero [29], 2017, Ukraine and United States | Examine whether social media provides LGBTQ youths with a safe space for identity exploration and expression | 14 to 17 (mean 16.3) | 19 | Lesbian (n=3), gay (n=8), bisexual (n=1), queer (n=1), unsure (n=3), and not straight (n=3) | Recruitment: flyers sent to LGBTQ organizations and Facebook; data collection: questionnaires and surveys; measures: Social Media Frequency Survey and Facebook Intensity Scale | Three-quarters of Facebook users never or rarely experienced cyberbullying and considered it a safe space for connecting and communicating with others. Over two-thirds of participants reported social media to be a comfortable environment compared with offline. | LGBTQ social media users felt safe to communicate and explore with peers on platforms such as Facebook. |

| Study, year, and country | Purpose | Age (years) | Sample size, N | LGBTQ[a] sample | Method | Findings | Summary |
|---|---|---|---|---|---|---|---|
| McConnell et al [51], 2018, United States (Chicago) | Examine the relationship between Facebook and LGBTQ youth identity management | 19 to 28 (mean 24.13, SD 1.64) | 199 | Identifying as male (n=77), identifying as female (n=108), transwomen (n=15), transmen (n=3), gay (n=69), lesbian (n=55), bisexual (n=49), heterosexual (n=10), and unsure (n=8) | Recruitment: LGBTQ youths from a longitudinal study; data collection: questionnaires and surveys; measures: adapted Outness Inventory | Participants were grouped into 4 categories of level of identity disclosure on Facebook: cluster 1 (high overall outness), cluster 2 (low overall outness), cluster 3 (less out to family), and cluster 4 (more out to family). Cluster 1 comprised 64% of the participants with high levels of disclosure among family, classmates or colleagues, and others. | LGBTQ youths felt that free self-expression on social media was complicated because of factors relating to identity disclosure. By investigating Facebook accounts, youths were mostly categorized as being of either low or high outness. Some would purposely censor their identity expression to avoid unintentional identity disclosures. |
| McConnell et al [52], 2017, United States (Chicago) | Examine Facebook use among LGBTQ youths, identity management methods, and effects of outness | Mean 24.02 (SD 1.65) | 204 | Transgender (n=24), gay (n=59), lesbian (n=49), bisexual (n=42), heterosexual (n=9), and unsure (n=5) | Recruitment: LGBTQ youths from a longitudinal study via email and flyers sent to LGBTQ organizations; data collection: questionnaires and surveys; measures: adapted Outness Inventory, Multidimensional Scale of Perceived Social Support, and Brief Symptom Inventory | Over 13% had multiple Facebook accounts, and >42% used privacy settings to limit viewable content for selected friends. Participants reported high outness offline and on Facebook, both positively correlated ($r$=0.72; $P$<.001). Facebook outness showed a high positive correlation ($r$=0.73) and the lowest correlation among friends ($r$=0.53). | Social media can act as a strategy for identity management, which some users find important. Some LGBTQ youths possessed multiple accounts or platforms where they could differ identity expression according to audience. |
| McInroy et al [30], 2019, Canada and United States | Explore LGBTQ engagement with web-based and offline communities, activities, and resources | 14 to 29 (mean 18.35, SD 3.64) | 4009 | LGBTQ+ (n=7986), heterosexual (n=58), and cisgender (n=2211) | Recruitment: from LGBTQ organizations and school groups; data collection: questionnaires and surveys; measures: 6-scale questionnaire on activeness, support, and safety in web-based and offline LGBTQ communities | LGBTQ participants would connect more with the LGBTQ community on the web (88%) compared with offline (69%). LGBTQ participants were more engaged (2-tailed $t_{4008}$=10.12; $P$<.001) and supported ($t_{4008}$=26.28; $P$<.001) and safer ($t_{4008}$=35.78; $P$<.001) on the web compared with offline. LGBTQ social media or blogs were used by 87% of the participants, and identity-specific web or YouTube series were used by 79% of the participants. | LGBTQ youths were likely to participate on the web with other LGBTQ people, including social media. Social media was reported to be a safer, more supportive, and more active option compared with offline. |

| Study, year, and country | Purpose | Age (years) | Sample size, N | LGBTQ[a] sample | Method | Findings | Summary |
|---|---|---|---|---|---|---|---|
| Pellicane et al [55], 2020, United States (Midwest) | Examine relationships between social media acceptance and hostility and their effects on mental health | Mean 19.87 | 387 | Heterosexual (n=326), bisexual (n=40), homosexual (n=7), and other (n=5) | Recruitment: undergraduate psychology students from an electronic database; data collection: questionnaires and surveys; measures: Center for Epidemiological Studies-Depression Scale, State-Trait Anxiety Inventory, and Social Media Experiences Questionnaire | There were significant associations between acceptance via social media and reduced symptoms of depression (β=−0.453; *P*<.001). Higher social media acceptance was also significantly associated with reduced anxiety symptoms (β=−0.343; *P*<.001). Conversely, hostility on social media was associated with increased symptoms of depression (β=.120; *P*=.19). | Social media has the benefit of acceptance and support for LGBTQ individuals and can help prevent or reduce anxiety and depression. This pattern was not reflected among the non-LGBTQ population in this study. |
| Twist et al [59], 2017, United States (Southwest) | Explore LGB experiences of monitoring web-based visibility and relationships | 18 to 41 (mean 24.67) | 61 | Bisexual (n=33) and same-sex orientated (n=28) | Recruitment: undergraduate students; data collection: questionnaires and surveys; measures: Lesbian, Gay, and Bisexual Identity Scale as well as the Ecological Elements Questionnaire, Family Adaptability and Cohesion Scale-IV, and Same-Sexting Practices and Questionnaire | Facebook had high levels of visibility regarding LGB identity, relationship disclosure (32%), gender identity (30%), and sexuality (31%). Almost half (49%) of the participants felt that partner outness on the web was immaterial. Most (70%) reported infrequent negative responses to web-based identity disclosure. | Most participants reported their sexual identity via social media primarily on Facebook. Most participants did not report negative interactions because of their identity disclosure on social media. |

[a]LGBTQ: lesbian, gay, bisexual, transgender, and queer.

[b]LGB: lesbian, gay, and bisexual.

## Quality Assessment

Overall, the included articles were moderate to high in quality and limited to descriptive study designs. Both CASP and NOS scales ranged from 0 to 10. Qualitative data were of high standard, with a mean of 8.7 (SD 0.8, range 7-10), whereas quantitative data were of moderate standard, with a mean of 6.2 (SD 1.5, range 4-9). Half of the qualitative studies (9/18, 50%) were limited by a lack of sample description [31,33,49,50,56], having predominately White samples [27,28,32] with higher socioeconomic backgrounds [32], or having a small sample size [34]. A total of 6% (1/18) of the qualitative studies had only 8 participants [34], but the sample size was not included in the CASP. The importance of sample sizes in qualitative research to data adequacy is an open question [61]. A total of 6% (1/18) of the studies collected data from 2004 to 2006 [33]. In total, 11% (2/18) of the studies had small transgender subpopulations (n≤4) [53,54]. A total of 11% (2/18) of the studies were limited because of their restrictive recruitment (ie, primarily from 1 source) [27,48]. A total of 6% (1/18) of the studies recruited participants from a gender diversity clinic requiring parental permission, thus introducing bias by selecting youths with more supportive parents [56]. Of the 11 quantitative studies, 7 (64%) used cross-sectional designs [29-31,45,47] with a low [43,44] to medium risk of bias [29,59]. A total of 9% (1/11) of the studies used a longitudinal design but provided no information on loss to follow-up [55]. Only 18% (2/11) of the studies described the assessment and management of confounding [43,44]. Of the 11 studies, 4 (36%) had generalizability concerns [29-31,48], 2 (18%) had insufficient descriptions of the sample [31,46], 2 (18%) used inadequately validated measures [29,46], and 1 (9%) had significant volunteer bias [30]. A total of 9% (1/11) of the studies were limited because of the small sample size (n=19) [29].

## Connecting With Other LGBTQ Youths on Social Media

### *Overview*

Qualitative (13/18, 72%), quantitative (3/11, 27%), and mixed methods (2/3, 67%) studies found that web-based environments were safe spaces for LGBTQ peer connection [27-31,33,34]. Qualitative data on LGBTQ youth social media use consisted of subthemes, including that anonymity is used to connect with peers, youths connect differently depending on the social media platform, and social media reduces feelings of isolation. LGBTQ youths commonly connected with peers via social media platforms [27,31,32,54,60]. All the included studies (26/26, 100%) were at risk of volunteer bias because of the nature of the target population and recruitment methods.

### Qualitative Studies

A total of 72% (13/18) of the qualitative studies explored narratives about LGBTQ youths' connection with peers via social media [27,28,32-34,49-51,53,54,56,57,60]. Instagram, Tumblr, Twitter, and YouTube were commonly used to connect, at times anonymously [31,49,51,54,56]. A total of 8% (1/13) of the studies, which found that Facebook policies limited anonymity, were weakened by the fact that they did not specify an analysis method or describe recruitment [49]. Tumblr was popular among LGBTQ youths, providing community connection, information, and support [31,49,54]. Participants reported that they ceased using Tumblr once it became "toxic" and negatively affected their mental health [31]. Instagram users were able to find and connect with others via hashtags (eg, #lesbian), although it should be noted that this study was limited by selection biases as recruitment occurred by inviting participants through Instagram comments [50]. LGBTQ youths could cease negative interactions (eg, block profiles) easily via social media if they felt uncomfortable talking to others [54].

LGBTQ youths also resorted to social media to connect with the LGBTQ community when there was a lack of offline opportunities [34,54]. Social media was a vital tool for those in rural and remote settings to connect with LGBTQ peers [33]. Youths reported reduced feelings of isolation and increased well-being when connecting with other LGBTQ youths [33,54,56,60]. LGBTQ youths could converse with LGBTQ peers anonymously and, as comfort increased, meet offline [33,49,54]. The study by Varjas et al [60] required parental permission, which limited their sample to those with generally supportive parents. Developing a web-based and offline connection with those who shared the same identities helped form emotional connections within the community and between individuals (eg, romantic relationships) [28,33,34,54]. These platforms also acted as a mechanism for LGBTQ youths to engage in sexual encounters on the web or offline [54]. LGBTQ youths were more likely to meet their web-based connections in person compared with non-LGBTQ peers [28]. Many LGBTQ youths turned to web-based spaces such as social media as their offline environment was unaccepting [28,54].

### Quantitative Studies

A total of 45% (5/11) of the studies investigated peer and group LGBTQ connections among youths [29-31,45,46]. A total of 20% (1/5) of these studies reported that 65% of 1304 LGBTQ Tumblr users in Australia used the platform to connect with other LGBTQ youths [31]. Only 3% of the participants used Tumblr to connect with friends; rather, it was specifically used to interact with strangers who shared their identities [31]. Social media was used to connect with others, including LGBTQ celebrities or groups that improved their sense of belonging and provided gratification [30,45,46]. A total of 20% (1/5) of the studies noted that approximately 80% of 6178 LGBTQ youths followed LGBTQ celebrities and communities [46]. In total, 20% (1/5) of the studies identified that mental health and well-being were positively affected by social media connection, but this study was limited because of its small sample size of 19 adolescents, mainly gay men (42%) [29].

## LGBTQ Identity Development Using Social Media

### Overview

All study designs explored this theme (qualitative: 16/18, 89%; quantitative: 4/11, 36%; mixed methods: 1/3, 33%). This theme explored LGBTQ youth identity development and management through the use of social media strategies for identity expression, accessing information, and censorship. Subthemes included the use of anonymity and privacy settings, sharing and validating identity development experiences, and the disclosure of identity. These strategies focused on methods to avoid conflict and protect well-being [27,34,51,57]. Healthy identity development can improve mental health and well-being among LGBTQ youths [43,44,49].

### Qualitative Studies

Nearly all qualitative studies (16/18, 89%) explored concepts of LGBTQ identity development via social media [27,28,32-34,47-51,53,54,56-58,60]. Studies noted from participant narratives that Facebook, Tumblr, and Twitter tended to be used more than other platforms for facilitating identity development [27,28,31,49,54,58]. LGBTQ youths found social media vital for identity development as it reduced the danger and stigma of meeting in person [32,58]. LGBTQ youths developed understanding and acceptance of and comfort with their identity through exposure to experiences of peers via forums, videos, and written blogs [32,33,53,54,56]. Social media allowed these individuals to explore their identities safely and access gender identity transition information [28,32,48,54,56,57]. A total of 6% (1/16) of these studies had a risk of bias because of the use of telephone interviews only and a small sample size (n=13) [57].

Facebook, Tumblr, and Twitter were commonly mentioned platforms that facilitated identity expression and exploration [27,31,49,58]. Many turned to Tumblr and Twitter to specifically express their LGBTQ identity rather than Facebook because of its restrictive policies and audiences (ie, changing the name in the profile's URL and limited identity options) [31,49,58]. Young people found that connecting with LGBTQ communities allowed them to share experiences, for example, medical information and surgery experiences for transgender youths [33,48,53,56,60]. Many appreciated sharing feelings and lived experiences, reporting that other LGBTQ individuals understood them better compared with non-LGBTQ people [33,50]. Narratives from participants included how social media can be a safe environment that facilitates healthy identity development because of privacy setting features imperative to LGBTQ youths' web-based engagement [27,32,47,58]. Privacy settings and "friending" practices provided them with the ability to choose their social network audience and, therefore, how they expressed their identity [27,47,50]. This permitted LGBTQ youths to manage disclosure experiences such as gradually disclosing one's identity or remaining undisclosed if preferred [27,47].

Disclosing identity on the web provided the user with time to consider and articulate how they would communicate their identity to their offline networks [33]. Social media platforms such as Facebook Messenger allow gender-diverse users to change their nicknames to suit their identity, which could aid in gradual identity disclosure [28,49]. For some, it was vital to remain undisclosed to avoid danger, relationship deterioration, and negative interactions [47]. Social media could offer identity disclosure without the expectations, danger, and pressure associated with offline networks [28,32]. LGBTQ youths could express their identity by sharing with their audience using subtle posts (eg, images of same-sex partners, pronouns, names, and relationship statuses) [27,32,47,54,58].

Others reported that, if sexual preferences were left empty on Facebook, the person was considered likely not heterosexual [34,47]. Less subtle displays of "outness" usually occurred by having highly expressive and visible profiles [47]. These actions required considerable contemplation of the potential repercussions and reactions of audiences [32,51]. Social media offered a way for LGBTQ youths to disclose their identity without reprisal from friends or family [32]. Social media distanced LGBTQ youths from heteronormative environments, homophobia, and transphobia that they may have experienced offline [32,49,51,53,56,57]. There were mixed views of the platforms' (ie, Facebook's) use of LGBTQ-specific categories, with some praising the understanding of their identity and others finding it restrictive (eg, interested in men or women and other pre-existing terminology) [27,48]. A study focusing on Instagram users found that using identity hashtags was a better way to connect with peers [50].

Qualitative studies noted that having multiple social media accounts permitted LGBTQ youths to express and explore identities with specific audiences with anonymity [27,31-33,47,49-51]. Family, religious groups, and work were commonly named as audiences with whom LGBTQ youths needed multiple accounts and self-censorship to manage [51,58]. Pressure was experienced as friends and family monitored LGBTQ youths' social media [34,49]. Accidental disclosure of an LGBTQ identity, most commonly by sharing with unintended audiences, was identified as a risk of social media for identity expression [47,49,51]. Preventative strategies, which are often successful, included separating audiences, deidentifying locations and names, and adjusting privacy settings [47,49,51]. These strategies assisted in managing exposure to marginalization and stigma [49]. Even when censoring identity on social media, other indicators such as likes, images, group memberships, and friends' posts and events could be displayed [34,47]. Constantly monitoring and censoring references to LGBTQ content to avoid negative interactions could be overwhelming and cause youths to conform to heteronormative expectations [51].

Being able to view and interact with others expressing similar LGBTQ identities was validating for youths [33,53]. Seeing other youths, including schoolmates, engaging in LGBTQ-orientated activities on social media allowed for further identity exploration and understanding [33]. This exposure to other LGBTQ youths helped affirm one's identity and prove that LGBTQ people exist (eg, "liking" posted LGBTQ content) [28,33,50,51,53,56,57]. Shared backgrounds were another important factor for identity affirmation among ethnic minorities and religious groups [33,57,58]. Social media may assist in identity clashes (ie, LGBTQ and Christian identities) that create difficulties in understanding, exploration, and transition [58].

### *Quantitative Studies*

A total of 55% (6/11) of the studies examined LGBTQ identity development and management [29,43,44,51,52,59]. Overall, an increased understanding of identity via social media was associated with improved well-being outcomes [43,46]. Social media was reported as a safer and more comfortable approach for identity exploration than offline alternatives [29]. A study among same-sex attracted youths in the United States noted that 63% of 61 participants had their identity deliberately disclosed on social media [59]. This study had significant generalizability issues as recruitment was restricted to undergraduate students of 1 degree at 1 university [59]. Identity exploration and well-being were associated with higher use of social media among lesbian, gay, and bisexual (LGB) youths compared with non-LGB youths attending a straight-gay alliance at a US high school [43]. Identity exploration via social media was associated with lower paranoia scores among American LGB youths [44]. However, the heavy use of social media for identity exploration had negative mental health consequences, increasing loneliness and sensitivity to emotional, physical, or social stimuli [44].

In a study of American LGBTQ youths, 13% of 181 participants had multiple Facebook accounts for identity exploration or expression [52]. Of this sample, 27% had publicly visible profiles, whereas 54% restricted their profiles to friends [52]. In total, 43% restricted what their friends could view on their Facebook profiles [52]. High levels of disclosure on Facebook were common, with 64% of 199 LGBTQ participants freely displaying their identity [51]. Another study found that 30% of LGBTQ youths disclosed on Facebook, significantly higher than on other platforms, including Tumblr (5%-9%) or Twitter (8%-13%) [44]. The level of identity disclosure related to the individual's willingness to express their identity [51]. LGBTQ youths who were not disclosed to their family were often highly engaged with and disclosed to their LGBTQ networks on the web compared with those who were disclosed to their family [51]. LGBTQ youths did not consider their partners not being disclosed on social media or offline as an issue for their relationship or satisfaction, with 23 of 61 participants reporting that it was "extremely unimportant" [59].

## Social Support on Social Media

### *Overview*

This final theme explored the support mechanisms that LGBTQ youths used via social media (qualitative: 5/18, 28%; quantitative: 4/11, 36%; mixed methods: 2/3, 67%). Subthemes in the qualitative data on LGBTQ youths using social media included seeking social support from peers and communicating information and experiences between peers. LGBTQ youths would find support by connecting with other LGBTQ people or groups and obtaining pertinent information [28,33,53,54,56]. Access to social support and information can be beneficial for mental health and well-being [46,55].

### Qualitative Studies

Almost half (7/18, 39%) of the qualitative studies explored social support among LGBTQ youths through social media [28-31,33,54,56]. Social support among LGBTQ youths was more commonly reported as occurring on the web compared with non-LGBTQ youths, whose offline networks were sufficient [28]. Social media connections were useful for seeking support during difficult times for young LGBTQ individuals [28,54,56]. Facebook was used to participate in LGBTQ groups where individuals could express emotions and seek support [29,54].

Web-based friends could provide support without geographical restrictions as participants communicated with others in different countries [54,56]. Posting on social media about mental or physical health concerns was not always a method to elicit social support but, rather, to simply be heard [31]. LGBTQ youths were able to interact with other or experienced LGBTQ community members for advice on dating, safety, sex, identity disclosure, and sexuality [28,33,54]. Social support via social media was highly convenient and could be obtained whenever required, even at short notice [54]. Transgender youths were able to seek specific support from other transgender individuals and share transition experiences [53,56]. Many transgender youths reported viewing YouTube videos as support for their transition as such information was generally inaccessible offline [53]. Transgender youths were able to access pertinent medical information and resources [53].

### Quantitative Studies

A total of 45% (5/11) of the studies examined social support through social media among LGBTQ youths [31,44-46,55]. Social media afforded LGBTQ youths support that they might not have achieved offline [46]. Ceglarek and Ward [44] reported that, among their 570 participants, there was strong evidence ($\beta=-0.27$; $P\leq.01$) that the use of social media for support was linked to reduced loneliness and evidence ($\beta=-0.21$; $P\leq.05$) of reduced paranoia among LGB youths from self-reported data. There was weak or no evidence ($P>.05$) that social media use among LGB youths reduced anxiety and depression [44]. Social media acceptance and support were associated with reduced symptoms of depression among 387 LGB youths ($\beta=.453$; $P<.001$) but had no significant effect on anxiety; however, this study did not describe loss to follow-up [55]. Social media use was associated with feelings of being loved or feeling stronger [45]. Although Tumblr was not a uniformly positive experience, 30% of 1304 surveyed LGBTQ youths reported it as a useful resource [31]. As age increased, the use of social media for social support and information among 6178 participants decreased significantly ($P<.001$) [46].

## Discussion

### Principal Findings

Similar systematic reviews have been published previously [62,63], but this is the first to explore how LGBTQ youths use social media and its impact on mental health and well-being. This systematic review explored how LGBTQ youths use social media and how it affects their peer connections, identity exploration, and social support. We found 26 studies—overall, the quality of the research was moderate and limited to observational studies. Most studies (16/26, 62%) were limited by a lack of follow-up, limited description of study confounders, restrictive sample eligibility limiting generalizability, and selection biases. Causality cannot be inferred from associations. With these limitations in mind, how LGBTQ youths use social media for connection and identity has been well explored. Both the positive and negative impacts of social media use among LGBTQ youths were identified.

We identified multiple ways in which social media has a positive impact on the mental health and well-being of LGBTQ youths. The studies showed a reduction in mental illness symptoms, including anxiety, depression, and paranoia [34,44,45,55,60]. Participant narratives identified decreased feelings of isolation and increased well-being when engaging in social media [33,54,56,60]. Social media was a significant source of social support for LGBTQ youths [28,33,54,56]. It was a setting where young people could control the expression of their sexual and gender identities to prevent or reduce exposure to stigma and discrimination [27,28,32-34,47-51,53-55]. Alleviating stressors among LGBTQ youths is associated with a reduced risk of poor mental health, including depression and suicidal ideation [43]. Other literature has found that developing networks and expressing LGBTQ identity safely (without stigma or discrimination) leads to reduced mental health problems, including anxiety, depression, addictive behaviors, and suicidal ideation [11,64].

We also identified the negative outcomes of social media use. Heavy social media use among LGBTQ youths was associated with increased feelings of loneliness and sensitivity [44]. Social media dependency was also linked to poorer academic performance, sleep deprivation, and mental health conditions [62,65]. Although social media could limit discrimination and stigma, LGBTQ youths are still at higher risk of web-based victimization [66], and other research has noted that social media can be a source of discrimination, including within web-based LGBTQ networks [26]. Mental health and well-being were negatively affected by social media structures and policies that did not accommodate changed identities [49]. A study found that chosen identity recognition may be associated with reduced mental illness symptoms [67]. Much of the wider literature identifies the effects of social media among LGBTQ youths as generally negative [52]. However, this review identifies both the positive and negative aspects of social media use among LGBTQ youths. Some platforms (eg, Grindr, Tinder, and Twitter) offer users nonbinary options, but displaying this information could also lead to safety issues [47,68]. Being “outed” on social media can target LGBTQ youths for physical harm and discriminatory comments, with the potential to affect their mental health and well-being [55]. Although a study exploring the use of hashtags only found positive and negative outcomes of identity exploration [50], another study found that hashtags can be used for other purposes, such as connecting peers that use drugs (eg, #highlife) [69].

We found that social media allowed LGBTQ youths to actively manage their identity, whereas non-LGBTQ youths did not demonstrate the same use of social media or the need to explicitly express their identity [27,34,50,51,58]. LGBTQ youths would actively manage their audiences by friending those of similar ages, limiting some of them via privacy settings, or removing friends [27,47]. These platforms' affordances made it possible for LGBTQ youths to connect with numerous other LGBTQ people and disclose their identity regardless of physical location [33,54,70]. In contrast, non-LGBTQ individuals reported sufficient support offline and did not add strangers to their social media [28].

This review noted that Facebook and Twitter had higher identity disclosure than other platforms and sexuality-specific dating apps [59]. Web-based sexual encounters usually occur via geosocial networking apps that allow for location sharing among LGBTQ and non-LGBTQ individuals (eg, Grindr, Tinder, and Bumble) [54,71,72]. Other social media platforms such as Facebook were also used for this purpose but less commonly [73]. Connections among LGBTQ individuals or communities can lead to romantic relationships, overcoming barriers such as fewer potential romantic partners and societal restrictions [11]. Dating same-sex partners was associated with improved mental health and self-esteem and reduced internalized homophobia compared with dating other sex partners [11]. The wider literature found that, for young people, negative aspects of social media included jealousy (eg, images of their partner with other people) or inaccurate social media depictions of relationships [74,75]. LGBTQ relationship portrayals and web-based engagement on social media may also affirm one's identity [28,33,50,51,53,56,57]. It is important to note that social media landscapes can change rapidly in platform popularity, as seen in recent years with the gradual decrease in Tumblr use and the rise of TikTok [76,77].

### *Limitations*

Only published peer-reviewed data (and no gray literature) were included. This systematic review was also limited as the sensitive nature of sexual and mental health meant that individual studies were at risk of reporting bias. There were very few studies (5/26, 19%) that investigated social media use influences on the mental health of LGBTQ youths. In addition, there was no uniform measure assessing mental health outcomes to determine the effect of social media use on mental health. Finally, owing to the ever-changing nature of social media and digital technology, these concepts may not capture current experiences.

### *Conclusions*

This review identified LGBTQ youths' uses of social media to connect with like-minded peers, manage their identity, and seek support. In the few studies that considered mental health outcomes (5/26, 19%), the use of social media appeared to be beneficial to the mental health and well-being of this group [11,34,44,55,60]. In this systematic review, we identified the various important beneficial roles of social media, but the findings were limited by weaknesses in the evidence base. This information may be useful for professionals (eg, educators, clinicians, and policy makers) working with LGBTQ youth to consider the appropriate use of social media in interventions as it provides an evidence base for the role of social media in the lives of LGBTQ youths. These findings help further understand how LGBTQ youths use social media and its positive and negative impacts on their mental health and well-being. Further research is required to provide stronger evidence of how social media is used for connectivity, identity, and support and determine causal links to mental health outcomes. We recommend larger, representative, and prospective research, including intervention evaluation, to better understand the potential of social media to support the health and well-being of marginalized LGBTQ young people. It is imperative that social media is understood and its beneficial use is supported to ensure improved outcomes.

## Acknowledgments

The authors would like to thank Kanchana Ekanayake, who helped develop the literature search necessary for this systematic review. This review did not receive any specific funding.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
PROSPERO registration.
[PDF File (Adobe PDF File), 1873 KB - jmir_v24i9e38449_app1.pdf ]

Multimedia Appendix 2
Search strategy.
[DOCX File , 15 KB - jmir_v24i9e38449_app2.docx ]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 25 KB - jmir_v24i9e38449_app3.docx ]

Multimedia Appendix 4

Summary of the included qualitative studies (N=18).
[DOCX File , 30 KB - jmir_v24i9e38449_app4.docx ]

---

## References

1. Household use of information technology. Australian Bureau of Statistics. 2018 Mar 28. URL: https://www.abs.gov.au/statistics/industry/technology-and-innovation/household-use-information-technology/latest-release [accessed 2021-04-10]
2. Fisher C, Waling A, Kerr L, Bellamy R, Ezer P, Mikolajczak G. 6th national survey of secondary students and sexual health. 6th National Survey of Australian Secondary Students and Sexual Health. 2019. URL: https://www.researchgate.net/publication/333812458_6th_National_Survey_of_Australian_Secondary_Students_and_Sexual_Health_2018 [accessed 2021-05-03]
3. Anderson M, Jiang J. Teens, social media and technology. Pew Research Center. 2018 May 31. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-05-04]
4. Trifiro BM, Gerson J. Social media usage patterns: research note regarding the lack of universal validated measures for active and passive use. Social Media Soc 2019 May 25;5(2):205630511984874. [doi: 10.1177/2056305119848743]
5. Subrahmanyam K, Reich SM, Waechter N, Espinoza G. Online and offline social networks: use of social networking sites by emerging adults. J Applied Development Psychol 2008 Nov;29(6):420-433. [doi: 10.1016/j.appdev.2008.07.003]
6. Spies Shapiro LA, Margolin G. Growing up wired: social networking sites and adolescent psychosocial development. Clin Child Fam Psychol Rev 2014 Mar;17(1):1-18 [FREE Full text] [doi: 10.1007/s10567-013-0135-1] [Medline: 23645343]
7. Pempek TA, Yermolayeva YA, Calvert SL. College students' social networking experiences on Facebook. J Applied Development Psychol 2009 May;30(3):227-238. [doi: 10.1016/j.appdev.2008.12.010]
8. Benamar L, Balagué C, Ghassany M. The identification and influence of social roles in a social media product community. J Comput Mediat Commun 2017;22(6):337-362. [doi: 10.1111/jcc4.12195]
9. Alhabash S, Ma M. A tale of four platforms: motivations and uses of Facebook, Twitter, Instagram, and snapchat among college students? Social Media Soc 2017 Feb 01;3(1):205630511769154. [doi: 10.1177/2056305117691544]
10. Frost DM, Meyer IH, Schwartz S. Social support networks among diverse sexual minority populations. Am J Orthopsychiatry 2016 Jan;86(1):91-102 [FREE Full text] [doi: 10.1037/ort0000117] [Medline: 26752447]
11. Russell ST, Fish JN. Mental health in Lesbian, Gay, Bisexual, and Transgender (LGBT) youth. Annu Rev Clin Psychol 2016 Mar 28;12(1):465-487 [FREE Full text] [doi: 10.1146/annurev-clinpsy-021815-093153] [Medline: 26772206]
12. Meyer IH. The elusive promise of LGBT equality. Am J Public Health 2016 Aug;106(8):1356-1358. [doi: 10.2105/ajph.2016.303221]
13. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697 [FREE Full text] [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
14. Higa D, Hoppe MJ, Lindhorst T, Mincer S, Beadnell B, Morrison DM, et al. Negative and positive factors associated with the well-being of lesbian, gay, bisexual, transgender, queer, and questioning (LGBTQ) youth. Youth Soc 2012 Jun 13;46(5):663-687. [doi: 10.1177/0044118x12449630]
15. Rothman EF, Sullivan M, Keyes S, Boehmer U. Parents' supportive reactions to sexual orientation disclosure associated with better health: results from a population-based survey of LGB adults in Massachusetts. J Homosex 2012 Feb;59(2):186-200 [FREE Full text] [doi: 10.1080/00918369.2012.648878] [Medline: 22335417]
16. Katz-Wise SL, Rosario M, Tsappis M. Lesbian, gay, bisexual, and transgender youth and family acceptance. Pediatr Clin North Am 2016 Dec;63(6):1011-1025 [FREE Full text] [doi: 10.1016/j.pcl.2016.07.005] [Medline: 27865331]
17. Harlow AF, Lundberg D, Raifman JR, Tan AS, Streed CG, Benjamin EJ, et al. Association of coming out as lesbian, gay, and bisexual+ and risk of cigarette smoking in a nationally representative sample of youth and young adults. JAMA Pediatr 2021 Jan 01;175(1):56-63 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.3565] [Medline: 33104174]
18. McConnell EA, Birkett MA, Mustanski B. Typologies of social support and associations with mental health outcomes among LGBT youth. LGBT Health 2015 Mar;2(1):55-61 [FREE Full text] [doi: 10.1089/lgbt.2014.0051] [Medline: 26790019]
19. Snapp SD, Watson RJ, Russell ST, Diaz RM, Ryan C. Social support networks for LGBT young adults: low cost strategies for positive adjustment. Fam Relat 2015 Jun 05;64(3):420-430. [doi: 10.1111/fare.12124]
20. Bowling J, Barker J, Gunn LH, Lace T. "It just feels right": perceptions of the effects of community connectedness among trans individuals. PLoS One 2020 Oct 5;15(10):e0240295 [FREE Full text] [doi: 10.1371/journal.pone.0240295] [Medline: 33017435]
21. Ceatha N, Mayock P, Campbell J, Noone C, Browne K. The power of recognition: a qualitative study of social connectedness and wellbeing through LGBT sporting, creative and social groups in Ireland. Int J Environ Res Public Health 2019 Sep 27;16(19):3636 [FREE Full text] [doi: 10.3390/ijerph16193636] [Medline: 31569733]
22. #Outlawed: "the love that dare not speak its name". Human Rights Watch. URL: http://internap.hrw.org/features/features/lgbt_laws/ [accessed 2022-01-29]
23. Owens ZD. Is it Facebook official? Coming out and passing strategies of young adult gay men on social media. J Homosex 2017 Jul 29;64(4):431-449. [doi: 10.1080/00918369.2016.1194112] [Medline: 27229129]

24. McNeill T. Sex education and the promotion of heteronormativity. Sexualities 2013 Dec 05;16(7):826-846. [doi: 10.1177/1363460713497216]
25. Leung H, Shek D, Leung E, Shek E. Development of contextually-relevant sexuality education: lessons from a comprehensive review of adolescent sexuality education across cultures. Int J Environ Res Public Health 2019 Feb 20;16(4):621 [FREE Full text] [doi: 10.3390/ijerph16040621] [Medline: 30791604]
26. Berger MN, Taba M, Marino JL, Lim MS, Cooper SC, Lewis L, et al. Social media's role in support networks among LGBTQ adolescents: a qualitative study. Sex Health 2021;18(5):421. [doi: 10.1071/sh21110]
27. Bates A, Hobman T, Bell BT. "let me do what I please with it . . . don't decide my identity for me": LGBTQ+ youth experiences of social media in narrative identity development. J Adolescent Res 2019 Nov 03;35(1):51-83. [doi: 10.1177/0743558419884700]
28. Hillier L, Mitchell KJ, Ybarra ML. The internet as a safety net: findings from a series of online focus groups with LGB and non-LGB young people in the United States. J LGBT Youth 2012 Jul;9(3):225-246. [doi: 10.1080/19361653.2012.684642]
29. Lucero L. Safe spaces in online places: social media and LGBTQ youth. Multicultural Educ Rev 2017 Apr 12;9(2):117-128. [doi: 10.1080/2005615x.2017.1313482]
30. McInroy LB, McCloskey RJ, Craig SL, Eaton AD. LGBTQ+ youths' community engagement and resource seeking online versus offline. J Technol Human Service 2019 Jun 01;37(4):315-333. [doi: 10.1080/15228835.2019.1617823]
31. Byron P, Robards B, Hanckel B, Vivienne S, Churchill B. "hey, I'm having these experiences": tumblr use and young people's queer (dis)connections. Int J Commun 2019;13 [FREE Full text]
32. Craig SL, McInroy L. You can form a part of yourself online: the influence of new media on identity development and coming out for LGBTQ youth. J Gay Lesbian Mental Health 2014 Jan 30;18(1):95-109. [doi: 10.1080/19359705.2013.777007]
33. Harper GW, Serrano PA, Bruce D, Bauermeister JA. The internet's multiple roles in facilitating the sexual orientation identity development of gay and bisexual male adolescents. Am J Mens Health 2016 Sep 23;10(5):359-376 [FREE Full text] [doi: 10.1177/1557988314566227] [Medline: 25585861]
34. Rubin JD, McClelland SI. 'Even though it's a small checkbox, it's a big deal': stresses and strains of managing sexual identity(s) on Facebook. Cult Health Sex 2015 Jan 14;17(4):512-526. [doi: 10.1080/13691058.2014.994229] [Medline: 25587894]
35. Adolescent health. World Health Organization. URL: https://www.who.int/southeastasia/health-topics/adolescent-health [accessed 2022-06-11]
36. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomized studies in meta-analysis. The Ottawa Hospital. 2000. URL: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [accessed 2021-02-21]
37. Herzog R, Álvarez-Pasquin MJ, Díaz C, Del Barrio JL, Estrada JM, Gil A. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. BMC Public Health 2013 Feb 19;13:154 [FREE Full text] [doi: 10.1186/1471-2458-13-154] [Medline: 23421987]
38. CASP checklist. CASP. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2021-02-21]
39. 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]
40. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
41. Hall WJ, Dawes HC, Plocek N. Sexual orientation identity development milestones among lesbian, gay, bisexual, and queer people: a systematic review and meta-analysis. Front Psychol 2021 Oct 21;12:753954 [FREE Full text] [doi: 10.3389/fpsyg.2021.753954] [Medline: 34777153]
42. Mental health: strengthening our response. World Health Organization. 2022 Jun 17. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response [accessed 2022-06-11]
43. Bond BJ, Figueroa-Caballero A. Exploring the relationship between computer-mediated communication, sexual identity commitment, and well-being among lesbian, gay, and bisexual adolescents. Commun Res Report 2016 Oct 14;33(4):288-294. [doi: 10.1080/08824096.2016.1224159]
44. Ceglarek PJ, Ward LM. A tool for help or harm? How associations between social networking use, social support, and mental health differ for sexual minority and heterosexual youth. Comput Human Behav 2016 Dec;65:201-209. [doi: 10.1016/j.chb.2016.07.051]
45. Chong E, Zhang Y, Mak W, Pang I. Social media as social capital of LGB individuals in Hong Kong: its relations with group membership, stigma, and mental well-being. Am J Community Psychol 2015 Mar;55(1-2):228-238. [doi: 10.1007/s10464-014-9699-2] [Medline: 25576016]
46. Craig SL, Eaton AD, McInroy LB, Leung VW, Krishnan S. Can social media participation enhance LGBTQ+ youth well-being? Development of the social media benefits scale. Social Media Soc 2021 Jan 23;7(1):205630512198893. [doi: 10.1177/2056305121988931]
47. Duguay S. "He has a way gayer Facebook than I do": investigating sexual identity disclosure and context collapse on a social networking site. New Media Soc 2014 Sep 04;18(6):891-907. [doi: 10.1177/1461444814549930]

48. Fox J, Ralston R. Queer identity online: informal learning and teaching experiences of LGBTQ individuals on social media. Comput Human Behav 2016 Dec;65:635-642. [doi: 10.1016/j.chb.2016.06.009]
49. Hanckel B, Vivienne S, Byron P, Robards B, Churchill B. ‘That’s not necessarily for them’: LGBTIQ+ young people, social media platform affordances and identity curation. Media Culture Soc 2019 May 13;41(8):1261-1278. [doi: 10.1177/0163443719846612]
50. Herrera AP. Theorizing the lesbian hashtag: identity, community, and the technological imperative to name the sexual self. J Lesbian Stud 2018 Jul 03;22(3):313-328. [doi: 10.1080/10894160.2018.1384263] [Medline: 29173131]
51. McConnell E, Néray B, Hogan B, Korpak A, Clifford A, Birkett M. "Everybody puts their whole life on Facebook": identity management and the online social networks of LGBTQ youth. Int J Environ Res Public Health 2018 May 26;15(6):1078 [FREE Full text] [doi: 10.3390/ijerph15061078] [Medline: 29861439]
52. McConnell EA, Clifford A, Korpak AK, Phillips G, Birkett M. Identity, victimization, and support: Facebook experiences and mental health among LGBTQ youth. Comput Human Behav 2017 Nov;76:237-244 [FREE Full text] [doi: 10.1016/j.chb.2017.07.026] [Medline: 29225412]
53. McInroy LB, Craig SL. Transgender representation in offline and online media: LGBTQ youth perspectives. J Human Behav Social Environ 2015 Apr 09;25(6):606-617. [doi: 10.1080/10911359.2014.995392]
54. Paceley MS, Goffnett J, Sanders L, Gadd-Nelson J. "Sometimes you get married on Facebook": the use of social media among nonmetropolitan sexual and gender minority youth. J Homosex 2022 Jan 02;69(1):41-60. [doi: 10.1080/00918369.2020.1813508] [Medline: 32875962]
55. Pellicane MJ, Cooks JA, Ciesla JA. Longitudinal effects of social media experiences on depression and anxiety in LGB+ and heterosexual young adults. J Gay Lesbian Mental Health 2020 Jun 26;25(1):68-93. [doi: 10.1080/19359705.2020.1776805]
56. Selkie E, Adkins V, Masters E, Bajpai A, Shumer D. Transgender adolescents' uses of social media for social support. J Adolesc Health 2020 Mar;66(3):275-280. [doi: 10.1016/j.jadohealth.2019.08.011] [Medline: 31690534]
57. Singh AA. Transgender youth of color and resilience: negotiating oppression and finding support. Sex Roles 2012 Mar 18;68(11-12):690-702. [doi: 10.1007/s11199-012-0149-z]
58. Taylor Y, Falconer E, Snowdon R. Queer youth, Facebook and faith: Facebook methodologies and online identities. New Media Soc 2014 Jul 24;16(7):1138-1153. [doi: 10.1177/1461444814544000]
59. Twist ML, Bergdall MK, Belous CK, Maier CA. Electronic visibility management of lesbian, gay, and bisexual identities and relationships in young adulthood. J Couple Relationship Ther 2017 Jul 25;16(4):271-285. [doi: 10.1080/15332691.2016.1238798]
60. Varjas K, Meyers J, Kiperman S, Howard A. Technology hurts? Lesbian, gay, and bisexual youth perspectives of technology and cyberbullying. J School Violence 2012 Oct 03;12(1):27-44. [doi: 10.1080/15388220.2012.731665]
61. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exercise Health 2019 Dec 26;13(2):201-216. [doi: 10.1080/2159676x.2019.1704846]
62. Tsitsika AK, Tzavela EC, Janikian M, Ólafsson K, Iordache A, Schoenmakers TM, et al. Online social networking in adolescence: patterns of use in six European countries and links with psychosocial functioning. J Adolesc Health 2014 Jul;55(1):141-147. [doi: 10.1016/j.jadohealth.2013.11.010] [Medline: 24618179]
63. Escobar-Viera CG, Whitfield DL, Wessel CB, Shensa A, Sidani JE, Brown AL, et al. For better or for worse? A systematic review of the evidence on social media use and depression among lesbian, gay, and bisexual minorities. JMIR Ment Health 2018 Jul 23;5(3):e10496 [FREE Full text] [doi: 10.2196/10496] [Medline: 30037786]
64. Collier KL, van Beusekom G, Bos HM, Sandfort TG. Sexual orientation and gender identity/expression related peer victimization in adolescence: a systematic review of associated psychosocial and health outcomes. J Sex Res 2013 Apr;50(3-4):299-317 [FREE Full text] [doi: 10.1080/00224499.2012.750639] [Medline: 23480074]
65. Wolniczak I, Cáceres-DelAguila JA, Palma-Ardiles G, Arroyo KJ, Solís-Visscher R, Paredes-Yauri S, et al. Association between Facebook dependence and poor sleep quality: a study in a sample of undergraduate students in Peru. PLoS One 2013;8(3):e59087 [FREE Full text] [doi: 10.1371/journal.pone.0059087] [Medline: 23554978]
66. Abreu RL, Kenny MC. Cyberbullying and LGBTQ youth: a systematic literature review and recommendations for prevention and intervention. J Child Adolesc Trauma 2018 Mar 24;11(1):81-97 [FREE Full text] [doi: 10.1007/s40653-017-0175-7] [Medline: 32318140]
67. Russell ST, Pollitt AM, Li G, Grossman AH. Chosen name use is linked to reduced depressive symptoms, suicidal ideation, and suicidal behavior among transgender youth. J Adolesc Health 2018 Oct;63(4):503-505 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.02.003] [Medline: 29609917]
68. Albury K, Dietzel C, Pym T, Vivienne S, Cook T. Not your unicorn: trans dating app users' negotiations of personal safety and sexual health. Health Sociol Rev 2021 Mar 22;30(1):72-86. [doi: 10.1080/14461242.2020.1851610] [Medline: 33622202]
69. Stevens RC, Brawner BM, Kranzler E, Giorgi S, Lazarus E, Abera M, et al. Exploring substance use tweets of youth in the United States: mixed methods study. JMIR Public Health Surveill 2020 Mar 26;6(1):e16191 [FREE Full text] [doi: 10.2196/16191] [Medline: 32213472]
70. Gray M. Out in the Country Youth, Media, and Queer Visibility in Rural America. New York: NYU Press; 2009.

71. Macapagal K, Moskowitz DA, Li DH, Carrión A, Bettin E, Fisher CB, et al. Hookup app use, sexual behavior, and sexual health among adolescent men who have sex with men in the United States. J Adolesc Health 2018 Jun;62(6):708-715 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.01.001] [Medline: 29784114]
72. Sumter SR, Vandenbosch L. Dating gone mobile: demographic and personality-based correlates of using smartphone-based dating applications among emerging adults. New Media Soc 2018 Oct 20;21(3):655-673. [doi: 10.1177/1461444818804773]
73. Krishnan A, Nguyen M, Giang LM, Ha TV, Bhadra M, Nguyen SM, et al. Finding sex partners through social media among men who have sex with men in Hanoi, Vietnam. J Community Health 2018 Feb 4;43(1):146-156. [doi: 10.1007/s10900-017-0397-5] [Medline: 28677027]
74. Utz S, Beukeboom C. The role of social network sites in romantic relationships: effects on jealousy and relationship happiness. J Comput Mediated Commun 2011;16(4):511-527. [doi: 10.1111/j.1083-6101.2011.01552.x]
75. Taba M, Lewis L, Cooper SC, Albury K, Chung KS, Lim M, et al. What adolescents think of relationship portrayals on social media: a qualitative study. Sex Health 2020;17(5):467. [doi: 10.1071/sh20056]
76. Montag C, Yang H, Elhai JD. On the psychology of TikTok use: a first glimpse from empirical findings. Front Public Health 2021 Mar 16;9:641673 [FREE Full text] [doi: 10.3389/fpubh.2021.641673] [Medline: 33816425]
77. Dixon S. Statista. 2022 Feb 8. URL: https://www.statista.com/topics/2463/tumblr/#dossierKeyfigures [accessed 2022-06-19]

## Abbreviations

**CASP:** Critical Appraisal Skills Programme
**LGB:** lesbian, gay, and bisexual
**LGBTQ:** lesbian, gay, bisexual, transgender, and queer
**NOS:** Newcastle-Ottawa Quality Assessment Scale
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

*Edited by A Mavragani; submitted 02.04.22; peer-reviewed by ASW Chan , J Mayoh, S Song, C Escobar-Viera, R Plackett; comments to author 31.05.22; revised version received 22.08.22; accepted 29.08.22; published 21.09.22.*

*Please cite as:*
*Berger MN, Taba M, Marino JL, Lim MSC, Skinner SR*
*Social Media Use and Health and Well-being of Lesbian, Gay, Bisexual, Transgender, and Queer Youth: Systematic Review*
*J Med Internet Res 2022;24(9):e38449*
*URL: https://www.jmir.org/2022/9/e38449*
*doi:10.2196/38449*
*PMID:36129741*

Review

# Sustainability of Weight Loss Through Smartphone Apps: Systematic Review and Meta-analysis on Anthropometric, Metabolic, and Dietary Outcomes

Han Shi Jocelyn Chew[1], PhD; Wee Ling Koh[2], BSc; Janelle Shaina Hui Yi Ng[2], MSc; Ker Kan Tan[2,3], PhD

[1]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[2]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[3]Department of Surgery, National University Hospital, Singapore, Singapore

**Corresponding Author:**
Han Shi Jocelyn Chew, PhD
Alice Lee Centre for Nursing Studies
Yong Loo Lin School of Medicine
National University of Singapore
Level 3, Clinical Research Centre, Block MD11
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 65168687
Email: jocelyn.chew.hs@nus.edu.sg

## Abstract

**Background:** Evidence on the long-term effects of weight management smartphone apps on various weight-related outcomes remains scarce.

**Objective:** In this review, we aimed to examine the effects of smartphone apps on anthropometric, metabolic, and dietary outcomes at various time points.

**Methods:** Articles published from database inception to March 10, 2022 were searched, from 7 databases (Embase, CINAHL, PubMed, PsycINFO, Cochrane Library, Scopus, and Web of Science) using forward and backward citation tracking. All randomized controlled trials that reported weight change as an outcome in adults with overweight and obesity were included. We performed separate meta-analyses using random effects models for weight, waist circumference, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, blood glucose level, blood pressure, and total energy intake per day. Methodological quality was assessed using the Cochrane Risk of Bias tool.

**Results:** Based on our meta-analyses, weight loss was sustained between 3 and 12 months, with a peak of 2.18 kg at 3 months that tapered down to 1.63 kg at 12 months. We did not find significant benefits of weight loss on the secondary outcomes examined, except for a slight improvement in systolic blood pressure at 3 months. Most of the included studies covered app-based interventions that comprised of components beyond food logging, such as real-time diet and exercise self-monitoring, personalized and remote progress tracking, timely feedback provision, smart devices that synchronized activity and weight data to smartphones, and libraries of diet and physical activity ideas.

**Conclusions:** Smartphone weight loss apps are effective in initiating and sustaining weight loss between 3 and 12 months, but their effects are minimal in their current states. Future studies could consider the various aspects of the socioecological model. Conversational and dialectic components that simulate health coaches could be useful to enhance user engagement and outcome effectiveness.

**Trial Registration:** International Prospective Register of Systematic Reviews (PROSPERO) CRD42022329197; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=329197

***(J Med Internet Res 2022;24(9):e40141)*** doi:10.2196/40141

**KEYWORDS**

smartphone app; mobile app; mobile health; mHealth; eHealth; weight management; weight loss; obesity; app; diet; eating; mobile phone

## Introduction

Obesity is a metabolic disorder characterized by an excessive accumulation of fat, which is well known to increase one's risk of cardiometabolic diseases [1], psychological stress [2], and improve overall quality of life [3]. A weight reduction of 5%-10% has been shown to lower the risk of cardiometabolic diseases by improving cardiometabolic measures such as high waist circumference, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), high blood glucose level, and high blood pressure [4]. However, weight loss through behavioral modification remains challenging because of the lack of self-regulation and motivation [5].

Common weight management strategies include lifestyle modifications (eg diet and exercise), medications, and surgery, of which the former remains the safest, most conservative, and most adopted option. Self-regulation strategies such as self-monitoring, goal setting, action planning, and relapse prevention (eg, coping planning, stress and emotion management, and dietary lapse prevention) have been shown to improve weight loss and weight loss maintenance [6]. Recent research has also reported benefits of using technology to help users to promote a state of calorie deficit through dietary self-regulation. However, the implementation of such self-regulation strategies and their long-term effects on weight loss maintenance remain unclear.

Several meta-analyses have examined the effectiveness of smartphone apps on weight loss in adults but with several limitations. In 2015, the first meta-analysis on the effectiveness of mobile phone apps reported a significant weight loss of 1.04 kg [7]. This meta-analysis was based on 12 articles searched from 3 electronic databases. Articles on people with diseases other than obesity and mobile interventions with only SMS text messaging were excluded [7]. Another study focused on the effects of mobile apps on weight loss in the Asian population and reported a small to moderate interventional effect on weight loss (Hedges $g=-0.26$) [8]. Lastly, a study on both children and adults reported a pooled interventional effect- of −1.07 kg on weight loss. However, these reviews included studies of various study designs, such as randomized controlled trials (RCTs), case-control studies, and quasi-experimental studies, which could have reduced the quality and certainty of evidence [7]. Moreover, these systematic reviews included studies on both people with normal and high BMI [7,9] and studies on people with and without diseases (eg, cardiovascular diseases) [8]. These limitations render the interventional effects on each unique population unclear.

To the best of our knowledge, none of these systematic reviews examined the effects of smartphone apps on anthropometric, metabolic, and dietary outcomes across multiple time points to examine the sustainability of interventional effects on these outcomes. Therefore, we aimed to examine the effects of smartphone apps on anthropometric, metabolic, and dietary outcomes.

## Methods

### Overview

We conducted this review according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [10] (Multimedia Appendix 1) and registered it with PROSPERO (International Prospective Register of Systematic Reviews; CRD42022329197). Two or more reviewers assessed the study selection and risk of bias (ROB). Interrater agreements were assessed using Cohen κ, where Cohen $\kappa=0.00-0.20$ indicates no agreement, Cohen $\kappa=0.21-0.39$ indicates weak agreement, Cohen $\kappa=0.40-0.59$ indicates minimal agreement, Cohen $\kappa=0.60-0.79$ indicates moderate agreement, Cohen $\kappa=0.80-0.90$ indicates strong agreement, and Cohen $\kappa>0.90$ indicates almost perfect agreement [11].

### Search Strategy

We conducted a systematic search through 7 databases (CINAHL, PsycINFO, PubMed, Scopus, Cochrane Library, and Web of Science) for relevant articles published from database inception to March 10, 2022. Keywords and Medical Subject Headings terms were first searched through PubMed and Embase to permute more keywords, namely, *smartphone application*, *phone application*, *mobile*, *app*, *m-health*, *mHealth*, *obesity*, *overweight*, *body weight*, *weight loss*, *weight reduction*, and *weight management* (Multimedia Appendix 2 provides the detailed database keyword search). Citations were screened using the EndNote software (Clarivate). Full-text articles were independently screened by 3 reviewers (HSJC, JSHYN, and WLK).

### Study Selection

Titles and abstracts were first screened by HSJC according to the eligibility criteria crafted using the population, intervention, comparison condition, outcomes, and study design framework. Full-text articles were screened independently by HSJC, JSHYN, and WLK. Discrepancies were resolved among the 3 reviewers. Articles were included if they (1) were about people with high BMI (≥25 kg/m$^2$ for Western populations and ≥23 kg/m$^2$ for Asian populations; (2) examined the effectiveness of a smartphone app; (3) examined interventional effect on at least weight loss as an outcome; (4) reported outcomes beyond baseline and after intervention; and (5) had an RCT study design. Articles were excluded if they (1) focused on people with medical conditions (excluding overweight and obesity), (2) were not empirical RCTs (eg, pilot studies and secondary analyses of RCTs, in which case the original RCT was retrieved), (3) examined smartphone apps that only provided SMS text messaging interventions such as SMS text message reminders, and (4) did not have an English version of the manuscript.

### Data Extraction

Data extraction was performed independently by HSJC and WLK according to an Excel spreadsheet template for information on the following criteria: authors, year of publication, country, sample size, sample characteristics, the cutoff BMI score for having overweight (kg/$m^2$), mean age, percentage of male participants, socioeconomic status, educational level, weight measure, baseline weight, baseline BMI, app components, control condition, intervention duration, follow-up time points, attrition rate by the time of analysis, any significant baseline differences between the participants retained and those lost to follow-up, missing data management, whether the protocol was registered, whether the study was funded, and study outcomes in terms of mean (SD), mean (SE), or mean difference (MD; 95% CIs). For data extraction from 3-armed RCTs, pooled intervention outcome data were used if both intervention arms comprised different apps. Otherwise, only the intervention arm with a smartphone app was extracted. For studies that reported a separate intervention arm with additional nonapp components, data were not extracted.

### Methodological Quality

The Cochrane ROB Tool was used to rate the articles' methodological quality as low, unclear, and high ROB based on 6 domains, namely random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, outcome data completeness, and selective reporting [12]. Ratings were performed independently by JSHYN and WLK and discrepancies were resolved through discussions with HSJC. We planned to assess the certainty of evidence using the Grading of Recommendations, Assessment, Development, and Evaluation approach but decided not to because of the limited number of studies available and a relatively high ROB, which would not have allowed us to derive a meaningful conclusion on the certainty of evidence.

### Data Analysis

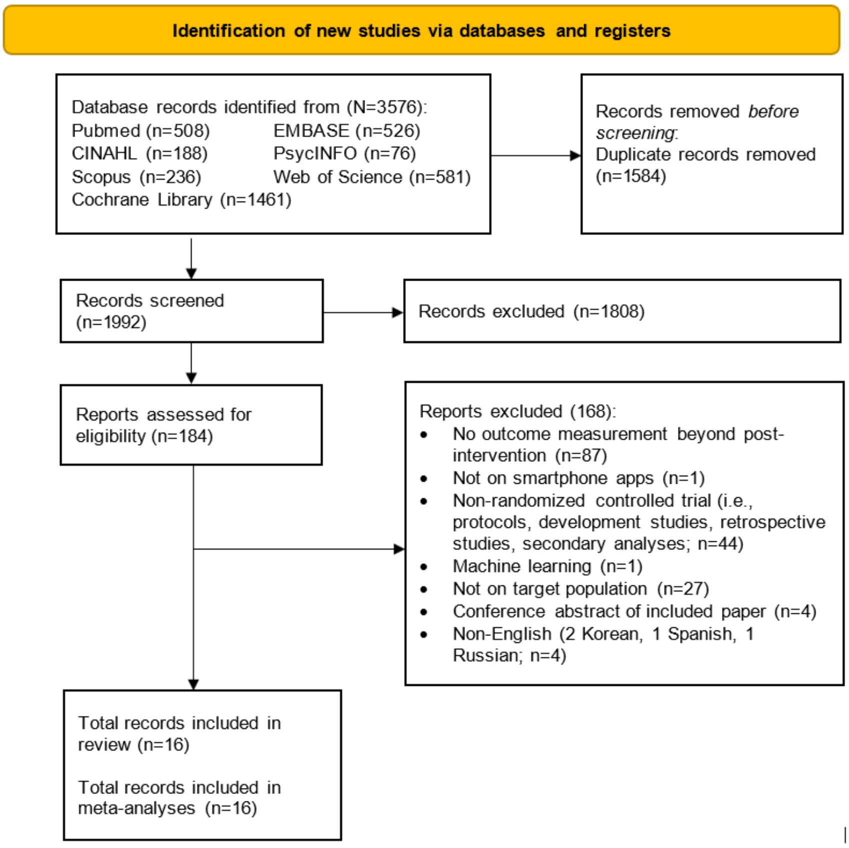
Sample variance estimates reported as SEs and CIs were converted to SD. The unit kJ/day was converted to kcal/day and lb was converted to kg. Weight, waist circumference, blood glucose, total energy intake, and blood pressure effect sizes were estimated using weighted MDs (WMDs), whereas HDL-C and LDL-C were estimated using standardized MDs (SMDs). SMDs were adjusted by using Hedges *g* to account for the small number of studies included in the meta-analyses. Some studies reported results from multiple time points beyond 6 months (ie, 9, 18, and 24 months). Therefore, results from the most reported time points were used for the meta-analyses to prevent bias arising from repeated measures. All meta-analyses were conducted using random effects models with generic inverse variance and adjusted using the Hartung-Knapp-Sidik-Jonkman method instead of the commonly used DerSimonian-Laird method, as it has been shown to produce fewer type I errors, especially for analyses with a small number of studies [13]. Between-study heterogeneity was estimated using the $\tau^2$ statistic and quantified using the $I^2$ statistic, where 25%, 50%, and 75% indicate a small, moderate, and large degree of heterogeneity, respectively [14]. Sensitivity analyses were performed to assess the individual effects of each study on the pooled effect size estimates, and publication bias was assessed using funnel plots and the Egger test, when possible. All statistical analyses were performed using R (version 4.1.3) [15].

## Results

### Overview

Of the 3576 articles retrieved from our systematic database search, 1584 (44.3%) duplicate articles were removed, resulting in 1992 (55.7%) titles and abstracts screened for eligibility. After excluding 90.8% (1808/1992) of articles based on the title and abstract screening, 9.2% (184/1992) full texts were assessed, of which 168 (91.3%) articles were excluded for reasons shown in Figure 1. We also searched through the references of the included articles, but no additional articles fulfilled the inclusion criteria. A total of 16 articles were included in the review, with 68 unique effect sizes included in the meta-analyses. The interrater agreement statistics for the inclusion of the article and overall ROB were Cohen $\kappa=0.77$; $P<.001$ and Cohen $\kappa=0.71$; $P=.003$, respectively, indicating moderate agreements.

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



## Study Characteristics

The 16 articles represented 2870 participants with overweight or obesity, with mean ages ranging from 22.7 years to 70.1 years, mean body weight ranging from 70.6 kg to 114.1 kg (2 studies did not report data on weight [16,17]), mean BMI ranging from 27.5 kg/m$^2$ to 36.2 kg/m$^2$, and the proportion of male participants ranging from 9.3% to 100%. Most studies (12/16, 75%) were conducted in the United States, except for 1 each from the United Kingdom [18], Australia [19], Japan [20], and China [21]. The attrition rates ranged from 0.5% to 46.6%. A total of 69% (11/16) of articles reported data that reflected the sample's socioeconomic status, and 81% (13/16) of articles reported data on the sample's education level. More details on the study characteristics are shown in Table 1, and additional information on socioeconomic status, education level, presence of group differences between participants retained and dropped out, protocol registration, and funding is shown in Multimedia Appendix 3 [16-31]. Most included studies (12/16, 75%) covered app-based interventions that were beyond purely food logging, such as real-time self-monitoring of diet and exercise, regular messages tailored according to user progress, timely feedback, smart devices that synchronized activity and weight data to smartphones, personalized progress reports, libraries of diet and physical activity ideas, and remote progress monitoring. Of these 16 articles, 4 (25%) included control conditions that provided app-based food logging [17,22-24]. The intervention duration ranged from 12 weeks to 24 months and the follow-up time points ranged from 8 weeks to 24 months. The intervention characteristics for each article are detailed in Multimedia Appendix 4 [16-31]. Most articles (10/16, 62%) were rated as having an unclear ROB, and 38% (6/16) of articles were rated as having a high ROB (Multimedia Appendix 5 [16-31]). Moreover, 38% (6/16) of articles were rated as having a high ROB for performance bias because of the difficulties in blinding both the interventionists and participants, which is common in such behavior-change studies (Multimedia Appendix 5 [16-31]). Owing to the varying outcome measurement time points, results were analyzed at <3 months, 3 months, 6 months, and >6 months whenever possible.

**Table 1.** Characteristics of the 16 included randomized controlled trial articles[a].

| Study | Trial arms (n); country | Sample size (n) | Age (years), mean | Proportion of males, % | Baseline weight (kg), mean; baseline BMI (kg/m$^2$), mean | Attrition rate[b], %; presence of group differences[c] | Missing data management; protocol registration; funding | Reported on SES[d]; educational level |
|---|---|---|---|---|---|---|---|---|
| Carter et al [18], 2013 | 3; United Kingdom | 128 | 41.9 | 22.7 | 96.9; 34.2 | 38.3; yes | ITT[e]; yes; yes | Yes; yes |
| Duncan [19], 2020 | 3; Australia | 116 | 44.5 | 29.3 | 90.7; 31.7 | 46.6; yes | ITT; yes; yes | Yes; yes |
| Dunn et al [17], 2019 | 2; United States | 43 | 42.4 | 9.3 | NS[f]; 34.5 | 30; NS | ITT; yes; yes | Yes; yes |
| Eisenhauer et al [22], 2021 | 2; United States | 80 | 54.2 | 100 | 114.1; 35.6 | 7.5; NS | NS; yes; yes | Yes; yes |
| Falkenhain et al [23], 2021 | 2; United States | 155 | 41 | 29 | 94.4; 33.5 | 25.2; nil | ITT; yes; yes | Yes; yes |
| Godino et al [30], 2016 | 2; United States | 404 | 22.7 | 29.7 | 80.7; 29 | 15.6; nil | ITT; yes; yes | Yes; yes |
| Johnston et al [26], 2013 | 2; United States | 292 | 46.5 | 10.2 | 90.1; 33 | 12; nil | ITT; no; yes | Yes; yes |
| Kurtzman et al [29], 2018 | 3; United States | 196 | 41.4 | 14.3 | 102.5; 36.2 | 4.5; NS | ITT; yes; yes | Yes; yes |
| Martin et al [25], 2015 | 2; United States | 40 | 44.4 | 17.5 | 80.3; 29.8 | 5.0; NS | ITT; yes; yes | NS; NS |
| Patel et al [24], 2019 | 3; United States | 105 | 42.7 | 16 | 89.6; 31.9 | 26.7; NS | ITT; yes; yes | Yes; yes |
| Rosas et al [31], 2020 | 2; United States | 192 | 50.2 | 38.2 | 87.1; 32.4 | 0.5; NS | PP[g]; yes; yes | Yes; yes |
| Ross et al [27], 2016 | 3; United States | 80 | 51.1 | 13.8 | 89.3; 33 | 10; NS | ITT; yes; yes | NS; NS |
| Spring et al [28], 2017 | 3; United States | 96 | 39.3 | 15.6 | 94.8; 34.6 | 13.5; yes | ITT; yes; yes | NS; Yes |
| Tanaka et al [20], 2018 | 2; Japan | 112 | 46.3 | 99.1 | 83.2; 28 | 27.7; NS | ITT; yes; yes | NS; NS |
| Turner-McGrievy et al [16], 2017 | 2; United States | 81 | 48.1 | 17.3 | NS; 33.4 | 25; nil | ITT; yes; yes | Yes; yes |
| Zhou et al [21], 2021 | 3; China | 750 | 70.1 | 46.1 | 70.6; 27.5 | 14.4; NS | NS; yes; yes | NS; Yes |

[a] Details of the group differences, protocol registration, funding, reports on socioeconomic status, and educational levels are presented in Multimedia Appendix 3 [16-31].

[b] Attrition rate by last outcome measurement time point.

[c] Group differences between participants who were retained and dropped out.

[d] SES: socioeconomic status.

[e] ITT: intention-to-treat.

[f] NS: not specified.

[g] PP: per-protocol analysis.

## Weight Loss

All 16 articles reported results on weight change. A total of 38% (6/16) of articles reported results of weight change at 3 months, of which 4 (67%) reported significant weight loss [18,20,21,25] and 2 (33%) reported otherwise [17,24]. Of these 4 articles, 2 (50%) reported results of 2 intervention arms with consistent findings on interventional effects on weight change and were analyzed as 4 studies in the meta-analysis [21,24]. The pooled WMD suggested a nonsignificant interventional effect on weight loss at <3 months (sample size, n=8; WMD=−1.15, 95% CI −3.02 to −0.72; $P$=.19; $I^2$=91.3%; Figure 2 and Table 2).

Of the 16 articles, 11 (69%) reported results of weight change at 3 months, of which 8 (50%) reported significant weight loss

[16,20,21,23,25-28] and 3 (19%) reported otherwise [22,24,29]. The pooled WMD for weight loss suggested statistically significant interventional effects at 3 months (n=11; WMD=−2.18, 95% CI −3.59 to −0.78; *P*=.006; $I^2$=87.3%). A subgroup analysis was also conducted, in which the heterogeneity between studies was not attributed to whether the respective control groups received an app-based intervention ($Q^2_1$=0.34; *P*=.56; Figure 3 and Table 2). No publication bias was detected based on the symmetry of the funnel plot and the Egger test (0.85; *t*=0.31; *P*=.76; Figure 4).

A total of 75% (12/16) of articles reported results of weight change at 6 months, of which 7 (44%) reported significant weight loss [16,18,22,23,26-28], and 5 (31%) reported otherwise [17,19,24,29,30]. The pooled WMD for weight loss suggested statistically significant interventional effects at 6 months (n=13; WMD=−2.15, 95% CI −3.25 to −1.05; *t*=−4.26; *P*=.001; $I^2$=52.4%; Figure 5 and Table 2). A subgroup analysis was also conducted, in which the heterogeneity between studies was not attributed to whether the respective control groups received an app-based intervention (*Q*=0.40; *P*=.53; Figure 5 and Table 2). No publication bias was detected based on the symmetry of the funnel plot and the Egger test (−0.74; *P*=.63; Figure 6).

In addition, 25% (4/16) of articles reported results of weight change at 12 months, of which 2 (50%) reported significant weight loss [30,31], and 2 (50%) reported otherwise [19,28]. Interventional effect was assessed at 9 months [29], 18 months [30], and 24 months [31]; only 6% (1/16) of articles reported significant weight loss at both 18 and 24 months [30]. The pooled WMD for weight loss suggested statistically significant interventional effects at 9 to 12 months (n=5; WMD=−1.63, 95% CI −2.99 to −0.26; *P*=.03; $I^2$=0%; Figure 7 and Table 2).

**Figure 2.** An illustration of the summary statistics of the intervention and control groups in each study included in the meta-analysis on the effect of smartphone weight loss apps on weight loss before 3 months. MD: mean difference.

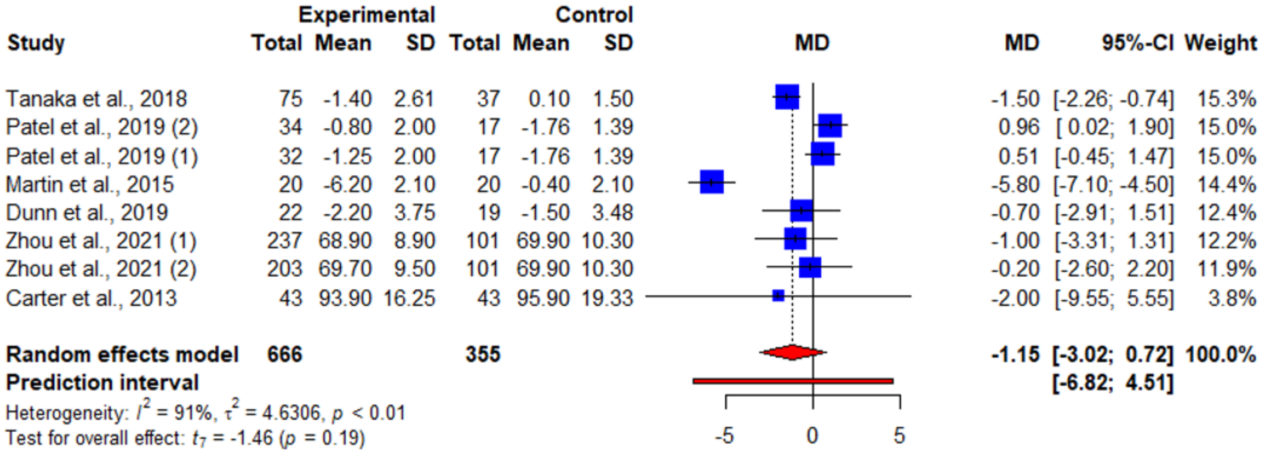


**Table 2.** A summary of meta-analyses results on each outcome at each time point analyzed.

| Time points | Sample size (n) | MD[a] or SMD[b] (95% CI) | *t* value | *P* value | $\tau^2$ statistic | $I^2$ (%) |
|---|---|---|---|---|---|---|
| **Weight (kg)** | | | | | | |
| <3 months | 8 | −1.15 (−3.02 to 0.72) | −1.46 | .19 | 4.63 | 91.3 |
| 3 months | 11 | −2.18 (−3.59 to −0.78) | −3.46 | .006[c] | 3.80 | 87.3 |
| 6 months | 13 | −2.15 (−3.25 to −1.05) | −4.26 | .001[d] | 2.02 | 52.4 |
| 9-12 months | 5 | −1.63 (−2.99 to −0.26) | −3.31 | .03[d] | 0.03 | 0 |
| **Waist circumference (cm)** | | | | | | |
| <3 months | 4 | −2.30 (−6.98 to −2.38) | −1.57 | .22 | 6.92 | 81.7 |
| 3 months | 4 | −3.85 (−9.31 to 1.60) | −2.25 | .11 | 10.28 | 88.7 |
| 6 months | 2 | −0.92 (−3.88 to 2.04) | −3.94 | .16 | 0 | 0 |
| 12 months | 3 | −1.19 (−3.80, to 1.43) | −1.95 | .19 | 0 | 0 |
| **HDL-C[e]** | | | | | | |
| 3 months | 2 | 0.01 (−0.15 to 0.17)[b] | 0.80 | .57 | 0 | 0 |
| **LDL-C[f]** | | | | | | |
| 3 months | 2 | −0.06 (−1.31 to 1.44)[b] | 0.58 | .66 | 0 | 0 |
| **Glycated hemoglobin ($HbA_{1c}$; %)** | | | | | | |
| 3-6 months | 3 | −0.22 (−1.03 to −0.6) | −1.14 | .37 | 0.07 | 67.1 |
| **Total energy intake per day[g]** | | | | | | |
| 6-12 months | 3 | −86.2 (−494.53 to 322.12) | −0.91 | .46 | 11,382 | 23.9 |
| **Systolic blood pressure (mm Hg)** | | | | | | |
| 3 months | 3 | −4.67 (−5.95 to −3.40) | −15.8 | .004[d] | 0 | 0 |
| 6 months | 2 | −0.28 (−15.6 to 15.03) | −0.23 | .85 | 1.10 | 21 |
| **Diastolic blood pressure (mm Hg)** | | | | | | |
| 3 months | 3 | −2.88 (−8.37 to 2.62) | −2.25 | .15 | 3.51 | 68.3 |
| 6 months | 2 | −0.65 (−1.56 to 0.26[c]) | −9.06 | .07 | 0 | 0 |

[a] MD: mean difference.

[b] SMD: standardized mean difference (adjusted with Hedges *g*).

[c] *P*<.05.

[d] *P*<.01.

[e] HDL-C: high-density lipoprotein cholesterol.

[f] LDL-C: low-density lipoprotein cholesterol.

[g] kJ converted to kcal.

**Figure 3.** An illustration of the summary statistics of the intervention and control groups in each study included in the meta-analysis on the effect of smartphone weight loss apps on weight loss at 3 months. The illustration also shows the subgroup analysis of the studies based on whether the control group received an app-based intervention. MD: mean difference.

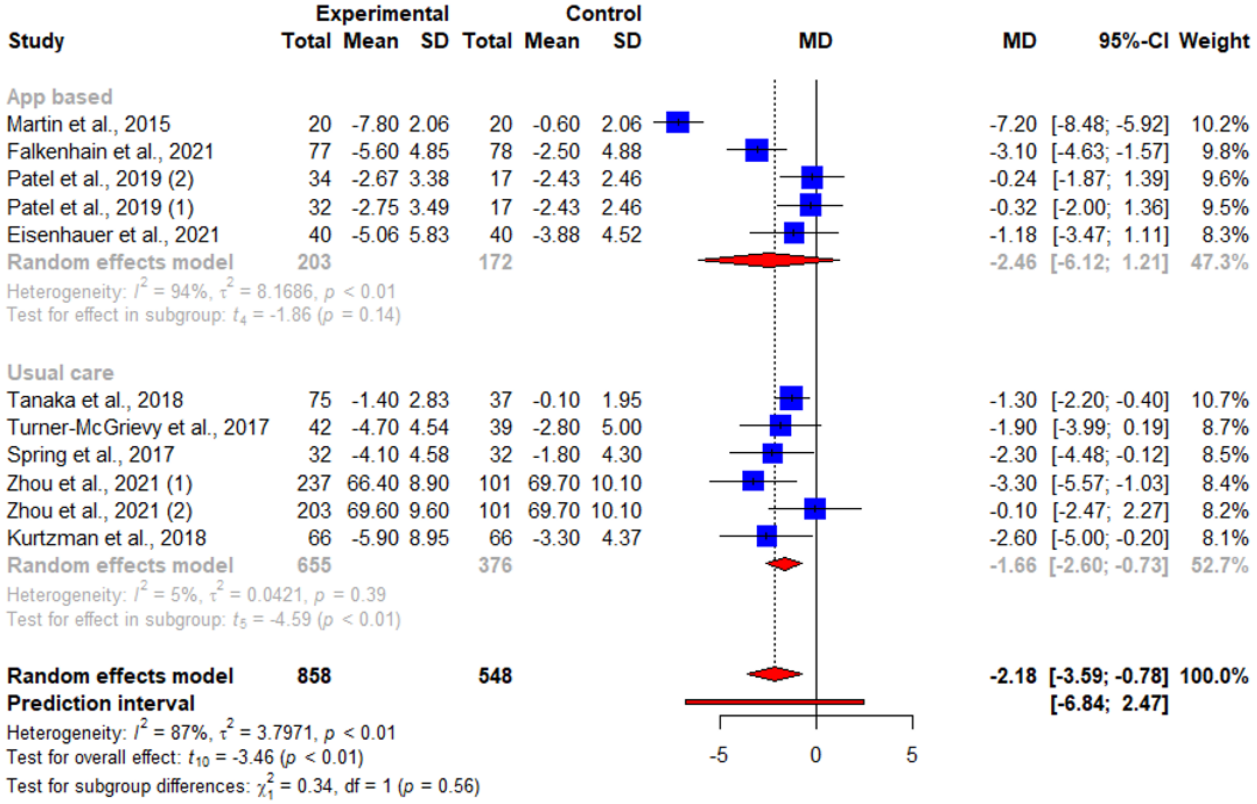


**Figure 4.** Funnel plot of symmetry for the included studies that reported the effects of smartphone weight loss apps on weight loss at 3 months.

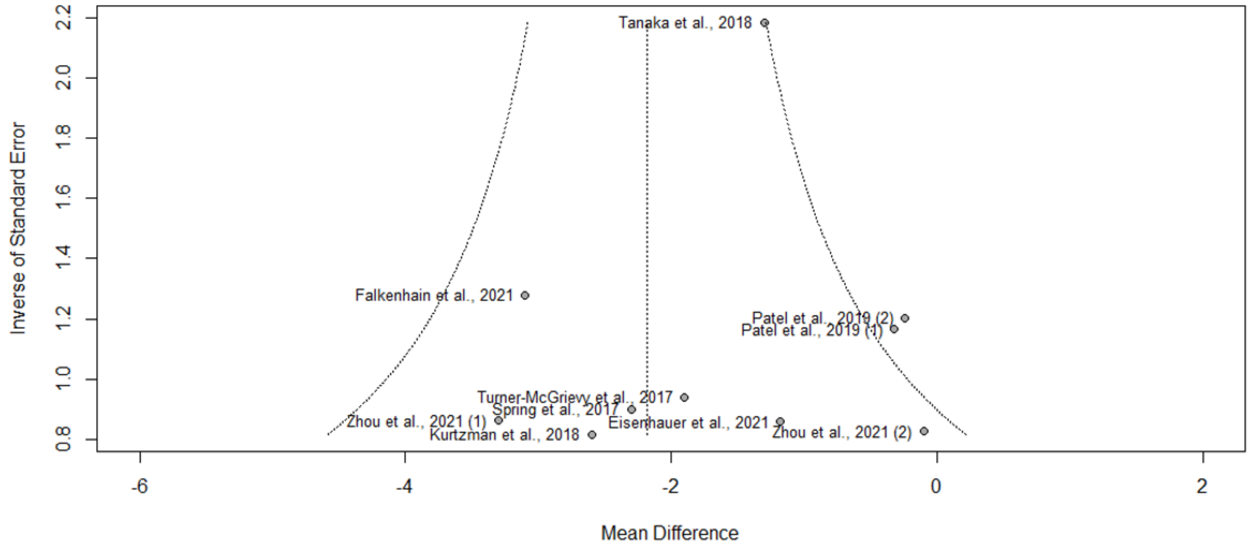
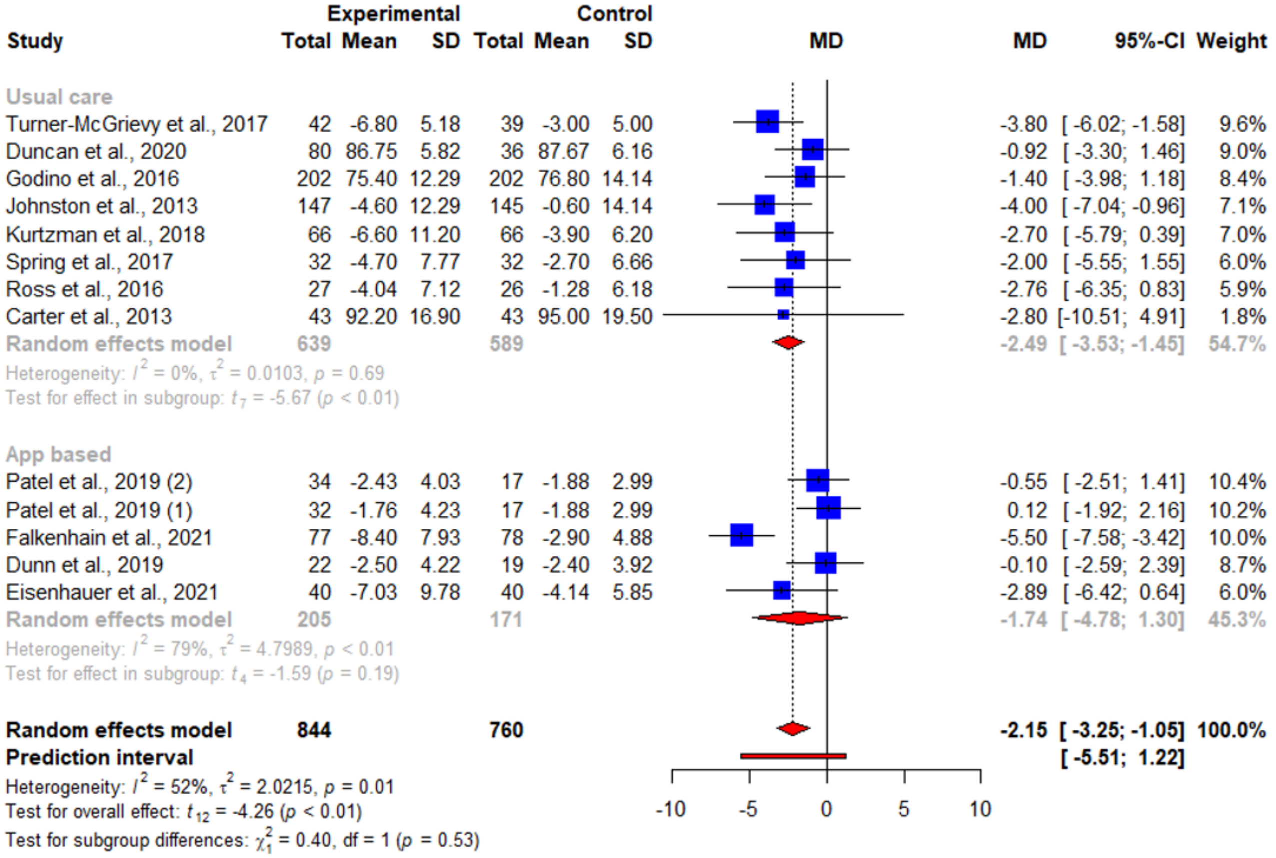


**Figure 5.** An illustration of the summary statistics of the intervention and control groups in each study included in the meta-analysis on the effect of smartphone weight loss apps on weight loss at 6 months. The illustration also shows the subgroup analysis of the studies based on whether the control group received an app-based intervention. MD: mean difference.



| Study | Experimental Total | Experimental Mean | Experimental SD | Control Total | Control Mean | Control SD | MD | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|---|---|
| Usual care | | | | | | | | | |
| Turner-McGrievy et al., 2017 | 42 | -6.80 | 5.18 | 39 | -3.00 | 5.00 | -3.80 | [-6.02; -1.58] | 9.6% |
| Duncan et al., 2020 | 80 | 86.75 | 5.82 | 36 | 87.67 | 6.16 | -0.92 | [-3.30; 1.46] | 9.0% |
| Godino et al., 2016 | 202 | 75.40 | 12.29 | 202 | 76.80 | 14.14 | -1.40 | [-3.98; 1.18] | 8.4% |
| Johnston et al., 2013 | 147 | -4.60 | 12.29 | 145 | -0.60 | 14.14 | -4.00 | [-7.04; -0.96] | 7.1% |
| Kurtzman et al., 2018 | 66 | -6.60 | 11.20 | 66 | -3.90 | 6.20 | -2.70 | [-5.79; 0.39] | 7.0% |
| Spring et al., 2017 | 32 | -4.70 | 7.77 | 32 | -2.70 | 6.66 | -2.00 | [-5.55; 1.55] | 6.0% |
| Ross et al., 2016 | 27 | -4.04 | 7.12 | 26 | -1.28 | 6.18 | -2.76 | [-6.35; 0.83] | 5.9% |
| Carter et al., 2013 | 43 | 92.20 | 16.90 | 43 | 95.00 | 19.50 | -2.80 | [-10.51; 4.91] | 1.8% |
| Random effects model | 639 | | | 589 | | | -2.49 | [-3.53; -1.45] | 54.7% |
| Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0.0103$, $p = 0.69$ | | | | | | | | | |
| Test for effect in subgroup: $t_7 = -5.67$ ($p < 0.01$) | | | | | | | | | |
| App based | | | | | | | | | |
| Patel et al., 2019 (2) | 34 | -2.43 | 4.03 | 17 | -1.88 | 2.99 | -0.55 | [-2.51; 1.41] | 10.4% |
| Patel et al., 2019 (1) | 32 | -1.76 | 4.23 | 17 | -1.88 | 2.99 | 0.12 | [-1.92; 2.16] | 10.2% |
| Falkenhain et al., 2021 | 77 | -8.40 | 7.93 | 78 | -2.90 | 4.88 | -5.50 | [-7.58; -3.42] | 10.0% |
| Dunn et al., 2019 | 22 | -2.50 | 4.22 | 19 | -2.40 | 3.92 | -0.10 | [-2.59; 2.39] | 8.7% |
| Eisenhauer et al., 2021 | 40 | -7.03 | 9.78 | 40 | -4.14 | 5.85 | -2.89 | [-6.42; 0.64] | 6.0% |
| Random effects model | 205 | | | 171 | | | -1.74 | [-4.78; 1.30] | 45.3% |
| Heterogeneity: $I^2 = 79\%$, $\tau^2 = 4.7989$, $p < 0.01$ | | | | | | | | | |
| Test for effect in subgroup: $t_4 = -1.59$ ($p = 0.19$) | | | | | | | | | |
| **Random effects model** | **844** | | | **760** | | | **-2.15** | **[-3.25; -1.05]** | **100.0%** |
| **Prediction interval** | | | | | | | | **[-5.51; 1.22]** | |

Heterogeneity: $I^2 = 52\%$, $\tau^2 = 2.0215$, $p = 0.01$

Test for overall effect: $t_{12} = -4.26$ ($p < 0.01$)

Test for subgroup differences: $\chi^2_1 = 0.40$, df = 1 ($p = 0.53$)

**Figure 6.** Funnel plot of symmetry for the included studies that reported the effects of smartphone weight loss apps on weight loss at 6 months.

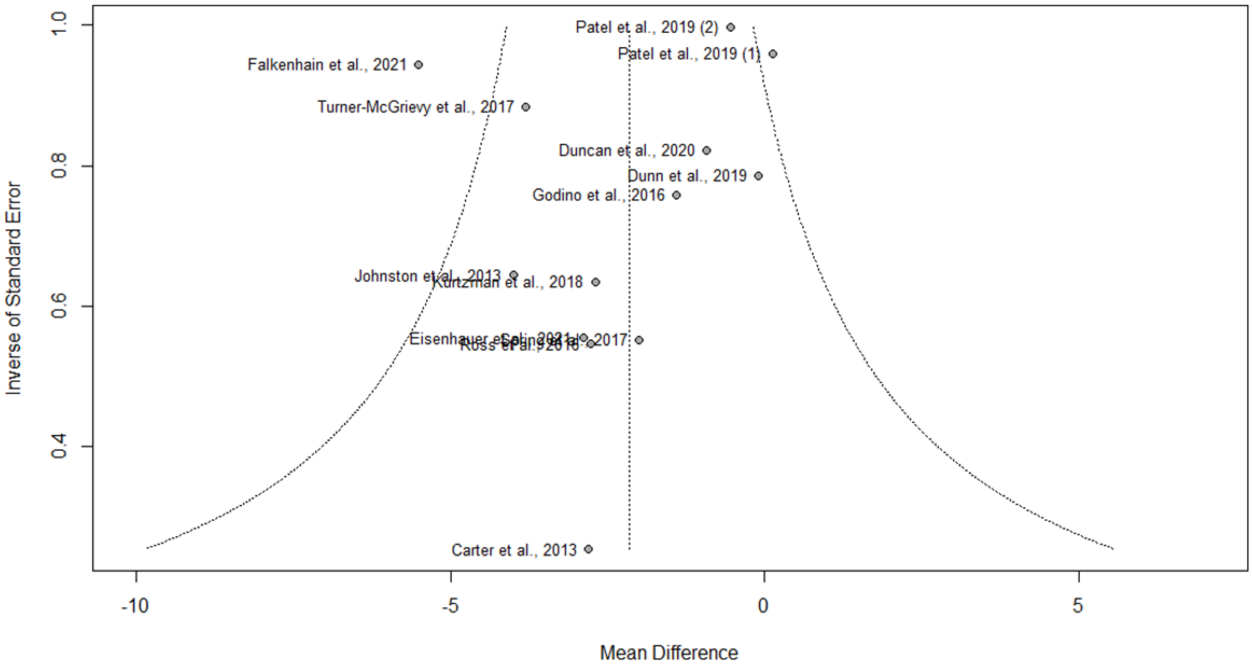
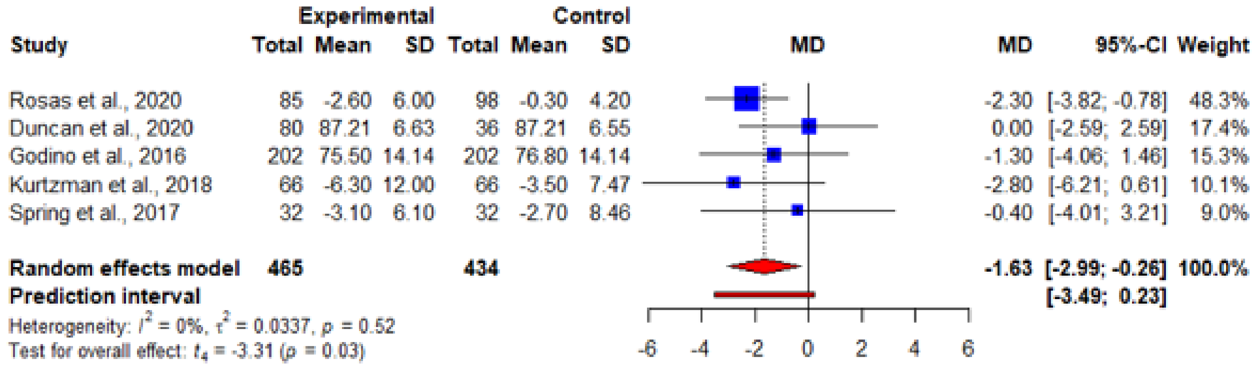


**Figure 7.** An illustration of the summary statistics of the intervention and control groups in each study included in the meta-analysis on the effect of smartphone weight loss apps on weight loss at 9 to 12 months. MD: mean difference.



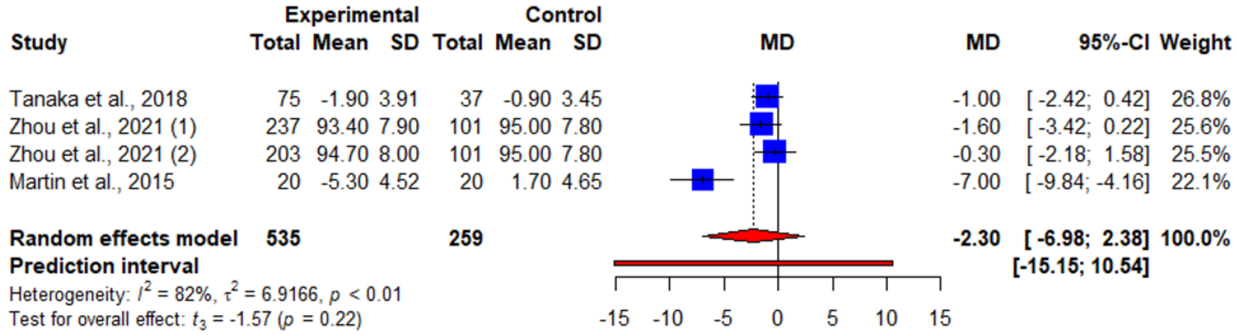
| Study | Experimental Total | Experimental Mean | Experimental SD | Control Total | Control Mean | Control SD | MD | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|---|---|
| Rosas et al., 2020 | 85 | -2.60 | 6.00 | 98 | -0.30 | 4.20 | -2.30 | [-3.82; -0.78] | 48.3% |
| Duncan et al., 2020 | 80 | 87.21 | 6.63 | 36 | 87.21 | 6.55 | 0.00 | [-2.59; 2.59] | 17.4% |
| Godino et al., 2016 | 202 | 75.50 | 14.14 | 202 | 76.80 | 14.14 | -1.30 | [-4.06; 1.46] | 15.3% |
| Kurtzman et al., 2018 | 66 | -6.30 | 12.00 | 66 | -3.50 | 7.47 | -2.80 | [-6.21; 0.61] | 10.1% |
| Spring et al., 2017 | 32 | -3.10 | 6.10 | 32 | -2.70 | 8.46 | -0.40 | [-4.01; 3.21] | 9.0% |
| **Random effects model** | **465** | | | **434** | | | **-1.63** | **[-2.99; -0.26]** | **100.0%** |
| **Prediction interval** | | | | | | | | **[-3.49; 0.23]** | |

Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0.0337$, $p = 0.52$
Test for overall effect: $t_4 = -3.31$ ($p = 0.03$)

## Waist Circumference

A total of 31% (5/16) of the articles measured interventional effects on waist circumference [19-21,25,30]. Of these 5 articles, 2 (40%) [21,25] reported a significant reduction in waist circumference at 3 months, whereas 1 (20%) reported otherwise [20]. Significant reductions in waist circumference were reported at 3 months [20,21,25] and 6 months [19,30], whereas different results were reported at 12 to 24 months [19,30]. No significant interventional effect was found on waist circumference beyond 4 months. No significant interventional effect was found through our meta-analyses for waist circumference before 3 months (Figure 8) and at 3 months (Figure 9), 6 months, and 12 months (Table 2).

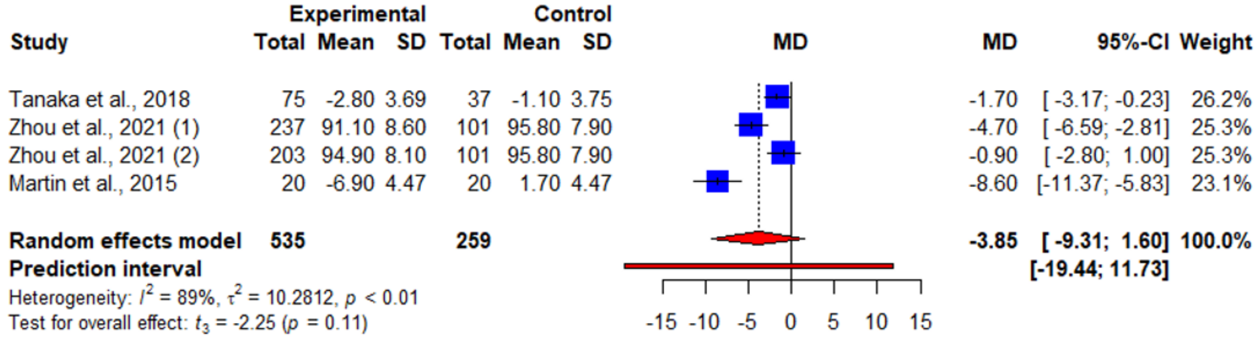
**Figure 8.** An illustration of the summary statistics of the intervention and control groups in each study included in the meta-analysis on the effect of smartphone weight loss apps on waist circumference before 3 months. MD: mean difference.



| Study | Experimental Total | Experimental Mean | Experimental SD | Control Total | Control Mean | Control SD | MD | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|---|---|
| Tanaka et al., 2018 | 75 | -1.90 | 3.91 | 37 | -0.90 | 3.45 | -1.00 | [-2.42; 0.42] | 26.8% |
| Zhou et al., 2021 (1) | 237 | 93.40 | 7.90 | 101 | 95.00 | 7.80 | -1.60 | [-3.42; 0.22] | 25.6% |
| Zhou et al., 2021 (2) | 203 | 94.70 | 8.00 | 101 | 95.00 | 7.80 | -0.30 | [-2.18; 1.58] | 25.5% |
| Martin et al., 2015 | 20 | -5.30 | 4.52 | 20 | 1.70 | 4.65 | -7.00 | [-9.84; -4.16] | 22.1% |
| **Random effects model** | **535** | | | **259** | | | **-2.30** | **[-6.98; 2.38]** | **100.0%** |
| **Prediction interval** | | | | | | | | **[-15.15; 10.54]** | |

Heterogeneity: $I^2 = 82\%$, $\tau^2 = 6.9166$, $p < 0.01$
Test for overall effect: $t_3 = -1.57$ ($p = 0.22$)

**Figure 9.** An illustration of the summary statistics of the intervention and control groups in each study included in the meta-analysis on the effect of smartphone weight loss apps on waist circumference at 3 months. MD: mean difference.



| Study | Experimental Total | Experimental Mean | Experimental SD | Control Total | Control Mean | Control SD | MD | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|---|---|
| Tanaka et al., 2018 | 75 | -2.80 | 3.69 | 37 | -1.10 | 3.75 | -1.70 | [-3.17; -0.23] | 26.2% |
| Zhou et al., 2021 (1) | 237 | 91.10 | 8.60 | 101 | 95.80 | 7.90 | -4.70 | [-6.59; -2.81] | 25.3% |
| Zhou et al., 2021 (2) | 203 | 94.90 | 8.10 | 101 | 95.80 | 7.90 | -0.90 | [-2.80; 1.00] | 25.3% |
| Martin et al., 2015 | 20 | -6.90 | 4.47 | 20 | 1.70 | 4.47 | -8.60 | [-11.37; -5.83] | 23.1% |
| **Random effects model** | **535** | | | **259** | | | **-3.85** | **[-9.31; 1.60]** | **100.0%** |
| **Prediction interval** | | | | | | | | **[-19.44; 11.73]** | |

Heterogeneity: $I^2 = 89\%$, $\tau^2 = 10.2812$, $p < 0.01$
Test for overall effect: $t_3 = -2.25$ ($p = 0.11$)

## Calorie Intake

A total of 25% (4/16) of articles measured interventional effects on waist circumference [16,19,23,31]. Of these 4 articles, 1 (25%) article reported a significant interventional effect on calorie intake per day at 3 and 6 months [16], whereas another article (1/4, 25%) reported otherwise [23]. In addition, 50% (2/4) of articles reported a significant reduction in calorie intake per day [19,23], and 50% (2/4) of articles reported no significant result at 12 months [19,31]. No significant interventional effect was found through our meta-analyses for total calorie intake per day between 6 and 12 months (Table 2).

## HDL-C, LDL-C, and HbA1c Levels

A total of 13% (2/16) of articles reported results on HDL-C and LDL-C, in which no significant change was reported at various time points [20,23]. In total, 19% (3/16) of articles reported results on $HbA_{1c}$, of which 13% (2/16) of articles reported a significant reduction at 3 months [20,23], and 13% (2/16) of articles reported results at 6 months but with contrasting findings [19,23]. According to our meta-analyses, no significant interventional effect was found for HDL-C, LDL-C, and $HbA_{1c}$ levels between 3 and 6 months (Table 2).

### Blood Pressure

A total of 19% (3/16) of articles reported results on blood pressure at various time points. A significant interventional effect was found for systolic blood pressure at 3 months (Cohen κ=3; WMD=−4.67, 95% CI −5.95 to −3.40; $t$=−15.8; $P$=.004; $I^2$=0%) but not at 6 months (Cohen κ=2; WMD=−0.28, 95% CI −15.6 to −15.03; $t$=−15.8; $P$=.004; $I^2$=0%). No significant interventional effect was found for diastolic blood pressure at 3 and 6 months (Table 2).

A summary of the outcomes reported in each study is reported in Multimedia Appendix 6 [16-31].

## Discussion

### Principal Findings

To the best of our knowledge, this is the first systematic review and meta-analysis that analyzed the effects of smartphone weight loss apps on weight loss and other anthropometric, metabolic, and dietary outcomes across various time points. Most articles reported results on weight loss at 3 and 6 months, but few reported findings on other anthropometric, metabolic, and dietary outcomes. Based on narrative syntheses and meta-analyses of evidence from the included studies, the findings showed that the use of smartphone apps for weight loss was generally minimal and unsustainable.

Based on our meta-analyses, weight loss was sustained between 3 and 12 months, with a peak of −2.18 kg at 3 months that tapered down with time to −1.63 at 12 months. This finding is similar to a prior systematic review that reported a significant weight loss of −1.99 kg and −2.8 kg at 3 and 6 months, respectively, in a population with and without overweight or obesity [32]. The slight difference in the time of peak weight loss could be related to the additional nonmobile elements implemented alongside smartphone apps, such as health coaching, which is well established to facilitate weight loss. However, weight loss of this magnitude may not be sufficient to reach a clinically significant reduction in cardiometabolic disease risk, assuming that a 5% decrease in weight (generally accepted rule-of-thumb for a clinically meaningful weight loss) for someone with a weight of 70.6 kg (lower value of the mean age range) translates to a weight loss of 3.53 kg [33]. Moreover, the actual effectiveness of such apps could be even lower considering that participants who stay throughout the interventions have a certain level of motivation to lose weight, potentially augmenting the results. Conversely, recent studies have shown improvements in outcomes such as the risk of developing type 2 diabetes, glucose tolerance, blood pressure, and triglycerides, even with a lower weight loss of ≥2% [34]. Nevertheless, the weight loss derived from using smartphones remains limited in their current state and may only be useful for people with slight overweight. Moreover, most results were derived from a Western population, who may have different preferences, engagement, and metabolic responses as compared to an Eastern population. Future studies may consider population characteristics in their app development.

The limited weight loss observed could be related to the behavioral components included in the weight loss apps examined, as 1 study reported that some form of an intensive health coaching alongside the use of smartphone apps improved weight loss [32]. This could be because of the immaturity of app-based coaches in matching up to the competencies of real-life health coaches in terms of relatability, usability, and trust [35]. However, another study reported weak associations between behavioral components and the usability and effectiveness of smartphone apps for weight loss, suggesting the unclear role of behavioral components in similar apps [36]. Given the established effectiveness of health coaching on weight loss, future studies could explore the integration of simulated health coaches—potentially the use of conversational agents to enhance the acceptability and engagement of similar apps while reducing the manpower needed [35].

Comparing the studies that reported mixed findings on weight loss at <3 months, the significant weight loss reported in the 4 studies could be attributed to the inclusion of personalized messages delivered by the app, dieticians, or coaches [18,20,21,25], which were not present in the remaining 2 studies that only included the food logging function and reported nonsignificant findings [17,24]. This is consistent with previous studies that reported higher engagement with, preference for, and weight loss outcomes in eHealth programs that included personalized recommendations [37]. Moreover, calorie counting has been described as troublesome and dislikable by users who prefer to have more motivational aspects in weight loss apps [38]. Therefore, future studies could consider replacing manual food logging with more intuitive and automatic methods. For example, food logging can be supported by food image recognition technology, self-directed goal setting, and progress monitoring, similar to most smartphone weight loss apps, and can be personalized and calibrated using medical and constantly updated lifestyle history [39].

By contrast, our meta-analysis results suggest significant effects of smartphone weight loss apps on weight loss from 3 to 12 months. This is consistent with existing systematic reviews [7-9], although the time points at which the analyzed data were retrieved were unclear. However, both our narrative synthesis and meta-analysis did not show the benefits of weight loss on secondary outcomes, except for a slight improvement in systolic blood pressure at 3 months. This is in contrast with previous systematic reviews on studies with follow-up time points of ≥2 years. With every reduction of 10 kg in body weight, a reduction of 5% to 10% in cholesterol levels has been reported alongside significant improvements in HDL-C and LDL-C levels [40]. Similarly, for every reduction of 10 kg in body weight, a reduction of 6.0 mm Hg and 4.6 mm Hg in systolic and diastolic blood pressure levels has been reported [41]. Weight loss has also been correlated with a reduction in fasting blood glucose levels in both diabetic and nondiabetic populations [42]. Weight loss without a change in metabolic outcomes defeats the purpose of weight loss, which should be to reduce the risk of noncommunicable lifestyle diseases (eg, coronary artery diseases and diabetes) through an improvement in metabolic outcomes. However, these discrepancies are more likely because of the small number of studies included in the meta-analyses of the secondary outcomes at various time points, which could have caused the analyses to be underpowered in detecting true effects, if present. Therefore, any inferences made based on the results of the secondary outcomes should be made with discretion. Moreover, compared with prior studies, a weight loss as large as 10 kg may be required for significant improvements in cholesterol and blood pressure to be detected. Future studies should consider examining a set of standard weight-related outcomes, such as the 7 outcomes examined in this study, over multiple time points to elucidate clearer findings.

Finally, our findings did not suggest a significant reduction in total calorie intake per day, potentially because of the small sample size. However, similar findings were reported in a large RCT that included patients with existing medical conditions [43]. The use of a smartphone app together with a smart activity band resulted in greater weight loss 12 months into the program, where both the intervention and control groups were found to have decreased their calorie intake comparably. This could be because of an increase in the duration of light physical activity, which could also explain our findings on interventional effectiveness on weight loss but not calorie intake.

In this age of widespread smartphone penetration and social media influence, people with overweight and obesity are often exposed to fad diets (eg, ketogenic diet, intermittent fasting, and Atkins diet) [44], and popular physical activities such as high-intensity interval training, cardio training, and resistance training. However, studies have shown limited effects and potential health issues (eg, malnutrition, dehydration, and acute injuries) with such diets [45] and unsupervised intensive physical activities [46]. As weight management is an integral aspect of population health, we should not be typecasting or identifying people with overweight and obesity for weight management but empowering people with the skills and resources to manage their weight independently. This is especially because of the global weight gain observed during the COVID-19 pandemic [47], which could be associated with varying levels of personal motivation [5], peer influence [48], and self-regulation [49]. Although many purchased home gym equipment during this period, the sustained use of such equipment without peer influence or health coaching may not have been high.

## Recommendations

Given the limited effectiveness of smartphone apps in their current state, future studies could consider the following recommendations. Further apps or app-related programs may consider the incorporation of the socioecological model to include more factors within the complexities of overweight and obesity [50]. These include food production, societal influences, food consumption, biology, individual psychology, individual activity, and activity environment [51]. Future apps may also consider including the component of health coaching as a continual source of motivation, discipline, and guidance. This could be in the form of a human coach, conversational agent, or other interactive embodiment that simulates a human coach [52].

## Strengths and Limitations

Each step of the systematic review process, including the study selection, data extraction, and methodological assessment, was independently performed by at least two reviewers. We also reported our findings in terms of MDs instead of SMDs to provide more clinically relevant metrics on the effects of smartphone apps on weight loss. However, there were some limitations to this study. First, no gray literature was searched, which could have precluded some relevant articles, but it ensured the rigor of this study by excluding non–peer-reviewed articles. Second, the sample size of articles included in the meta-analyses was small, rendering any conclusions prone to inaccuracies and biases. Therefore, discretion is needed when readers draw inferences and conclusions based on our findings. Third, the heterogeneity between the studies included in each outcome meta-analysis was high, ranging from 21% to 91.3%. This suggests a certain degree of inaccuracy in our conclusions. Finally, most included studies were rated as having an unclear or high ROB, indicating the presence of an inherent ROB in our findings. Nevertheless, we appended additional sociodemographic information such as the socioeconomic and educational profile of each included study for readers to make better judgments of the findings presented in this paper.

## Conclusions

The use of smartphone weight loss apps at the current stage may not be sufficient to produce clinically meaningful health outcomes. Future studies could consider tackling more influencing factors of weight management at every level of the socioecological model to empower population weight management. Future studies could consider including conversational and dialectic component that simulates a health coach and provides personalized progress monitoring and feedback to enhance user engagement and outcome effectiveness.

---

## Acknowledgments

The authors thank Nicole Nagadarshini for her help with part of the data organization. This research is supported by the National University Health System Internal Grant Funding under NUHSRO/2021/063/RO5+6/FMPCHSRG-Mar21/01. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author or authors and do not reflect the views of the funding agency.

## Authors' Contributions

HSJC was involved conceptualization; data curation; formal analysis; investigation; methodology; software; supervision; validation; visualization; and writing, reviewing, and editing the original draft. WLK was involved in data curation, formal analysis, investigation, software, validation, and writing (review and editing). JSHYN was involved in the formal analysis, investigation, software, validation, and writing (review and editing). KKT was involved in conceptualization, resources, supervision, validation, and writing (review and editing).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 31 KB - jmir_v24i9e40141_app1.docx ]

Multimedia Appendix 2
Search strategy.
[DOCX File , 14 KB - jmir_v24i9e40141_app2.docx ]

Multimedia Appendix 3
Extended study characteristics of the 16 included articles.
[DOCX File , 19 KB - jmir_v24i9e40141_app3.docx ]

Multimedia Appendix 4
Intervention characteristics of the 16 included articles.
[DOCX File , 22 KB - jmir_v24i9e40141_app4.docx ]

Multimedia Appendix 5
Methodological quality assessment of 16 included articles using the Cochrane Risk of Bias tool.
[DOCX File , 14 KB - jmir_v24i9e40141_app5.docx ]

Multimedia Appendix 6
Summary of the outcomes reported in each study.
[DOCX File , 18 KB - jmir_v24i9e40141_app6.docx ]

## References

1. Miranda JJ, Barrientos-Gutiérrez T, Corvalan C, Hyder AA, Lazo-Porras M, Oni T, et al. Understanding the rise of cardiometabolic diseases in low- and middle-income countries. Nat Med 2019 Nov;25(11):1667-1679. [doi: 10.1038/s41591-019-0644-7] [Medline: 31700182]
2. Spinosa J, Christiansen P, Dickson JM, Lorenzetti V, Hardman CA. From socioeconomic disadvantage to obesity: the mediating role of psychological distress and emotional eating. Obesity (Silver Spring) 2019 Apr;27(4):559-564 [FREE Full text] [doi: 10.1002/oby.22402] [Medline: 30821100]
3. Chu DT, Minh Nguyet NT, Nga VT, Thai Lien NV, Vo DD, Lien N, et al. An update on obesity: mental consequences and psychological interventions. Diabetes Metab Syndr 2019;13(1):155-160. [doi: 10.1016/j.dsx.2018.07.015] [Medline: 30641689]
4. Han TS, Lean ME. A clinical perspective of obesity, metabolic syndrome and cardiovascular disease. JRSM Cardiovasc Dis 2016 Feb 25;5:2048004016633371 [FREE Full text] [doi: 10.1177/2048004016633371] [Medline: 26998259]
5. Chew HS, Gao Y, Shabbir A, Lim SL, Geetha K, Kim G, et al. Personal motivation, self-regulation barriers and strategies for weight loss in people with overweight and obesity: a thematic framework analysis. Public Health Nutr 2022 Sep;25(9):2426-2435. [doi: 10.1017/S136898002200043X] [Medline: 35190011]
6. Stubbs RJ, Duarte C, O'Driscoll R, Turicchi J, Michalowska J. Developing evidence-based behavioural strategies to overcome physiological resistance to weight loss in the general population. Proc Nutr Soc 2019 Nov;78(4):576-589. [doi: 10.1017/S0029665119001083] [Medline: 31670628]
7. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]

8. Ang SM, Chen J, Liew JH, Johal J, Dan YY, Allman-Farinelli M, et al. Efficacy of interventions that incorporate mobile apps in facilitating weight loss and health behavior change in the Asian population: systematic review and meta-analysis. J Med Internet Res 2021 Nov 16;23(11):e28185 [FREE Full text] [doi: 10.2196/28185] [Medline: 34783674]
9. Islam MM, Poly TN, Walther BA, Jack Li YC. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039 [FREE Full text] [doi: 10.2196/17039] [Medline: 32706724]
10. 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. Int J Surg 2021 Apr;88:105906. [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
11. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
12. Higgins JP, 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 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
13. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 2014 Feb 18;14:25 [FREE Full text] [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
14. Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I 2 is not an absolute measure of heterogeneity. Res Synth Methods 2017 Mar;8(1):5-18. [doi: 10.1002/jrsm.1230] [Medline: 28058794]
15. RStudio Team. RStudio: Integrated Development Environment for R. RStudio. 2022. URL: http://www.rstudio.com/ [accessed 2022-09-12]
16. Turner-McGrievy GM, Wilcox S, Boutté A, Hutto BE, Singletary C, Muth ER, et al. The dietary intervention to enhance tracking with mobile devices (DIET mobile) study: a 6-month randomized weight loss trial. Obesity (Silver Spring) 2017 Aug;25(8):1336-1342 [FREE Full text] [doi: 10.1002/oby.21889] [Medline: 28600833]
17. Dunn CG, Turner-McGrievy GM, Wilcox S, Hutto B. Dietary self-monitoring through calorie tracking but not through a digital photography app is associated with significant weight loss: the 2SMART pilot study-a 6-month randomized trial. J Acad Nutr Diet 2019 Sep;119(9):1525-1532. [doi: 10.1016/j.jand.2019.03.013] [Medline: 31155474]
18. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
19. Duncan MJ, Fenton S, Brown WJ, Collins CE, Glozier N, Kolt GS, et al. Efficacy of a multi-component m-health weight-loss intervention in overweight and obese adults: a randomised controlled trial. Int J Environ Res Public Health 2020 Aug 26;17(17):6200 [FREE Full text] [doi: 10.3390/ijerph17176200] [Medline: 32859100]
20. Tanaka K, Sasai H, Wakaba K, Murakami S, Ueda M, Yamagata F, et al. Professional dietary coaching within a group chat using a smartphone application for weight loss: a randomized controlled trial. J Multidiscip Healthc 2018 Jul 16;11:339-347 [FREE Full text] [doi: 10.2147/JMDH.S165422] [Medline: 30038502]
21. Zhou M, Zhang N, Zhang Y, Yan X, Li M, Guo W, et al. Effect of mobile-based lifestyle intervention on weight loss among the overweight and obese elderly population in China: a randomized controlled trial. Int J Environ Res Public Health 2021 Aug 21;18(16):8825 [FREE Full text] [doi: 10.3390/ijerph18168825] [Medline: 34444573]
22. Eisenhauer CM, Brito F, Kupzyk K, Yoder A, Almeida F, Beller RJ, et al. Mobile health assisted self-monitoring is acceptable for supporting weight loss in rural men: a pragmatic randomized controlled feasibility trial. BMC Public Health 2021 Aug 18;21(1):1568 [FREE Full text] [doi: 10.1186/s12889-021-11618-7] [Medline: 34407782]
23. Falkenhain K, Locke SR, Lowe DA, Reitsma NJ, Lee T, Singer J, et al. Keyto app and device versus WW app on weight loss and metabolic risk in adults with overweight or obesity: a randomized trial. Obesity (Silver Spring) 2021 Oct;29(10):1606-1614 [FREE Full text] [doi: 10.1002/oby.23242] [Medline: 34124856]
24. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing self-monitoring strategies for weight loss in a smartphone app: randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e12209 [FREE Full text] [doi: 10.2196/12209] [Medline: 30816851]
25. Martin CK, Miller AC, Thomas DM, Champagne CM, Han H, Church T. Efficacy of SmartLoss, a smartphone-based weight loss intervention: results from a randomized controlled trial. Obesity (Silver Spring) 2015 May;23(5):935-942 [FREE Full text] [doi: 10.1002/oby.21063] [Medline: 25919921]
26. Johnston CA, Rost S, Miller-Kovach K, Moreno JP, Foreyt JP. A randomized controlled trial of a community-based behavioral counseling program. Am J Med 2013 Dec;126(12):1143.e19-1143.e24. [doi: 10.1016/j.amjmed.2013.04.025] [Medline: 24135513]
27. Ross KM, Wing RR. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: a randomized pilot study. Obesity (Silver Spring) 2016 Aug;24(8):1653-1659 [FREE Full text] [doi: 10.1002/oby.21536] [Medline: 27367614]
28. Spring B, Pellegrini CA, Pfammatter A, Duncan JM, Pictor A, McFadden HG, et al. Effects of an abbreviated obesity intervention supported by mobile technology: the ENGAGED randomized clinical trial. Obesity (Silver Spring) 2017 Jul;25(7):1191-1198 [FREE Full text] [doi: 10.1002/oby.21842] [Medline: 28494136]

29. Kurtzman GW, Day SC, Small DS, Lynch M, Zhu J, Wang W, et al. Social incentives and gamification to promote weight loss: the LOSE IT randomized, controlled trial. J Gen Intern Med 2018 Oct;33(10):1669-1675 [FREE Full text] [doi: 10.1007/s11606-018-4552-1] [Medline: 30003481]
30. Godino JG, Merchant G, Norman GJ, Donohue MC, Marshall SJ, Fowler JH, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. Lancet Diabetes Endocrinol 2016 Sep;4(9):747-755 [FREE Full text] [doi: 10.1016/S2213-8587(16)30105-X] [Medline: 27426247]
31. Rosas LG, Lv N, Xiao L, Lewis MA, Venditti EM, Zavella P, et al. Effect of a culturally adapted behavioral intervention for Latino adults on weight loss over 2 years: a randomized clinical trial. JAMA Netw Open 2020 Dec 01;3(12):e2027744 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.27744] [Medline: 33337491]
32. Antoun J, Itani H, Alarab N, Elsehmawy A. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 08;10(4):e35479 [FREE Full text] [doi: 10.2196/35479] [Medline: 35394443]
33. Magkos F, Fraterrigo G, Yoshino J, Luecking C, Kirbach K, Kelly SC, et al. Effects of moderate and subsequent progressive weight loss on metabolic function and adipose tissue biology in humans with obesity. Cell Metab 2016 Apr 12;23(4):591-601 [FREE Full text] [doi: 10.1016/j.cmet.2016.02.005] [Medline: 26916363]
34. Ryan DH, Yockey SR. Weight loss and improvement in comorbidity: differences at 5%, 10%, 15%, and over. Curr Obes Rep 2017 Jun;6(2):187-194 [FREE Full text] [doi: 10.1007/s13679-017-0262-y] [Medline: 28455679]
35. Chew HS, Achananuparp P. Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. J Med Internet Res 2022 Jan 14;24(1):e32939 [FREE Full text] [doi: 10.2196/32939] [Medline: 35029538]
36. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
37. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019 Feb 5;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]
38. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
39. Chew HS, Ang WH, Lau Y. The potential of artificial intelligence in enhancing adult weight loss: a scoping review. Public Health Nutr 2021 Jun;24(8):1993-2020 [FREE Full text] [doi: 10.1017/S1368980021000598] [Medline: 33592164]
40. Poobalan A, Aucott L, Smith WC, Avenell A, Jung R, Broom J, et al. Effects of weight loss in overweight/obese individuals and long-term lipid outcomes--a systematic review. Obes Rev 2004 Feb;5(1):43-50. [doi: 10.1111/j.1467-789x.2004.00127.x] [Medline: 14969506]
41. Aucott L, Poobalan A, Smith WC, Avenell A, Jung R, Broom J. Effects of weight loss in overweight/obese individuals and long-term hypertension outcomes: a systematic review. Hypertension 2005 Jun;45(6):1035-1041. [doi: 10.1161/01.HYP.0000165680.59733.d4] [Medline: 15897373]
42. Stanford J, Kaiser M, Ablah E, Dong F, Paull-Forney B, Early J. The effect of weight loss on fasting blood sugars and hemoglobin A1c in overweight and obese diabetics and non-diabetics. J Diabetes Mellitus 2012 Feb;2(1):126-130. [doi: 10.4236/jdm.2012.21021]
43. Lugones-Sanchez C, Recio-Rodriguez JI, Agudo-Conde C, Repiso-Gento I, G Adalia E, Ramirez-Manent JI, EVIDENT 3 Investigators. Long-term effectiveness of a smartphone app combined with a smart band on weight loss, physical activity, and caloric intake in a population with overweight and obesity (Evident 3 study): randomized controlled trial. J Med Internet Res 2022 Feb 01;24(2):e30416 [FREE Full text] [doi: 10.2196/30416] [Medline: 35103609]
44. Spadine M, Patterson MS. Social influence on fad diet use: a systematic literature review. Nutr Health 2022 Sep;28(3):369-388. [doi: 10.1177/02601060211072370] [Medline: 35023412]
45. Khawandanah J, Tewfik I. Fad diets: lifestyle promises and health challenges. J Food Res 2016 Nov 15;5(6):80-94. [doi: 10.5539/jfr.v5n6p80]
46. Rynecki ND, Siracuse BL, Ippolito JA, Beebe KS. Injuries sustained during high intensity interval training: are modern fitness trends contributing to increased injury rates? J Sports Med Phys Fitness 2019 Jul;59(7):1206-1212. [doi: 10.23736/S0022-4707.19.09407-6] [Medline: 30758171]
47. Chew HS, Lopez V. Global impact of COVID-19 on weight and weight-related behaviors in the adult population: a scoping review. Int J Environ Res Public Health 2021 Feb 15;18(4):1876 [FREE Full text] [doi: 10.3390/ijerph18041876] [Medline: 33671943]
48. Leahey TM, Kumar R, Weinberg BM, Wing RR. Teammates and social influence affect weight loss outcomes in a team-based weight loss competition. Obesity (Silver Spring) 2012 Jul;20(7):1413-1418 [FREE Full text] [doi: 10.1038/oby.2012.18] [Medline: 22310234]
49. Dassen FC, Houben K, Allom V, Jansen A. Self-regulation and obesity: the role of executive function and delay discounting in the prediction of weight loss. J Behav Med 2018 Dec;41(6):806-818 [FREE Full text] [doi: 10.1007/s10865-018-9940-9] [Medline: 29802535]

50. Lang S, Gibson S, Ng KW, Truby H. Understanding children and young people's experiences pursuing weight loss maintenance using the socio-ecological Model: a qualitative systematic literature review. Obes Rev 2021 May;22(5):e13172. [doi: 10.1111/obr.13172] [Medline: 33331090]
51. Blundell JE, Baker JL, Boyland E, Blaak E, Charzewska J, de Henauw S, et al. Variations in the prevalence of obesity among European countries, and a consideration of possible causes. Obes Facts 2017;10(1):25-37 [FREE Full text] [doi: 10.1159/000455952] [Medline: 28190010]
52. Chew HS. The use of artificial intelligence-based conversational agents (Chatbots) for weight loss: scoping review and practical recommendations. JMIR Med Inform 2022 Apr 13;10(4):e32578 [FREE Full text] [doi: 10.2196/32578] [Medline: 35416791]

## Abbreviations

**HDL-C:** high-density lipoprotein cholesterol
**LDL-C:** low-density lipoprotein cholesterol
**MD:** mean difference
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PROSPERO:** International Prospective Register of Systematic Reviews
**RCT:** randomized controlled trial
**ROB:** risk of bias
**SMD:** standardized mean difference
**WMD:** weighted mean difference

*Edited by R Kukafka; submitted 08.06.22; peer-reviewed by C Jacob, B Nievas Soriano; comments to author 13.07.22; revised version received 14.07.22; accepted 18.07.22; published 21.09.22.*

*Please cite as:*
*Chew HSJ, Koh WL, Ng JSHY, Tan KK*
*Sustainability of Weight Loss Through Smartphone Apps: Systematic Review and Meta-analysis on Anthropometric, Metabolic, and Dietary Outcomes*
*J Med Internet Res 2022;24(9):e40141*
*URL: https://www.jmir.org/2022/9/e40141*
*doi:10.2196/40141*
*PMID:36129739*

Review

# Digital Health Interventions for Depression and Anxiety Among People With Chronic Conditions: Scoping Review

Amika Shah[1,2], BSc, MScCH; Neesha Hussain-Shamsy[1,2], BA, MHS; Gillian Strudwick[1,3], RN, PhD; Sanjeev Sockalingam[3,4,5], MHPE, MD; Robert P Nolan[5,6,7], CPsych, PhD; Emily Seto[1,2], 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 Global eHealth Innovation, University Health Network, Toronto, ON, Canada
[3]Centre for Addiction and Mental Health, Toronto, ON, Canada
[4]University Health Network, Toronto, ON, Canada
[5]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[6]Cardiac eHealth, Toronto General Hospital, University Health Network, Toronto, ON, Canada
[7]Institute of Medical Science, University of Toronto, Toronto, ON, Canada

**Corresponding Author:**
Amika Shah, BSc, MScCH
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College St.
Toronto, ON, M5T 3M6
Canada
Phone: 1 416 978 4326
Email: amika.shah@mail.utoronto.ca

## Abstract

**Background:** Chronic conditions are characterized by their long duration (≥1 year), need for ongoing medical attention, and limitations in activities of daily living. These can often co-occur with depression and anxiety as common and detrimental comorbidities among the growing population living with chronic conditions. Digital health interventions (DHIs) hold promise in overcoming barriers to accessing mental health support for these individuals; however, the design and implementation of DHIs for depression and anxiety in people with chronic conditions are yet to be explored.

**Objective:** This study aimed to explore what is known in the literature regarding DHIs for the prevention, detection, or treatment of depression and anxiety among people with chronic conditions.

**Methods:** A scoping review of the literature was conducted using the Arksey and O'Malley framework. Searches of the literature published in 5 databases between 1990 and 2019 were conducted in April 2019 and updated in March 2021. To be included, studies must have described a DHI tested with, or designed for, the prevention, detection, or treatment of depression or anxiety in people with common chronic conditions (arthritis, asthma, diabetes mellitus, heart disease, chronic obstructive pulmonary disease, cancer, stroke, and Alzheimer disease or dementia). Studies were independently screened by 2 reviewers against the inclusion and exclusion criteria. Both quantitative and qualitative data were extracted, charted, and synthesized to provide a descriptive summary of the trends and considerations for future research.

**Results:** Database searches yielded 11,422 articles across the initial and updated searches, 53 (0.46%) of which were included in this review. DHIs predominantly sought to provide treatment (44/53, 83%), followed by detection (5/53, 9%) and prevention (4/53, 8%). Most DHIs were focused on depression (36/53, 68%), guided (32/53, 60%), tailored to chronic physical conditions (19/53, 36%), and delivered through web-based platforms (20/53, 38%). Only 2 studies described the implementation of a DHI.

**Conclusions:** As a growing research area, DHIs offer the potential to address the gap in care for depression and anxiety among people with chronic conditions; however, their implementation in standard care is scarce. Although stepped care has been identified as a promising model to implement efficacious DHIs, few studies have investigated the use of DHIs for depression and anxiety among chronic conditions using such models. In developing stepped care, we outlined DHI tailoring, guidance, and intensity as key considerations that require further research.

***(J Med Internet Res 2022;24(9):e38030)*** doi:10.2196/38030

**KEYWORDS**

depression; anxiety; multiple chronic conditions; chronic disease; mental health; psychiatry; digital health; eHealth; telehealth; mobile health; mHealth; telemedicine

## *Introduction*

### Background

Chronic conditions often co-occur despite an emphasis on their singular occurrence in health interventions, research, and care [1,2]. Chronic conditions persist for long durations (≥1 year), require ongoing medical attention, and limit activities of daily living [3]. In Canada, national data have identified several common chronic conditions, including arthritis, asthma, diabetes mellitus, heart disease, chronic obstructive pulmonary disease, cancer, stroke, mood and anxiety disorders, and Alzheimer disease and related dementias [4]. When poorly managed, these conditions have been associated with negative outcomes, such as poor quality of life [5], increased health care use [6], and higher costs [7].

Concomitant mood and anxiety disorders, such as depression and anxiety, are of particular concern because of their high prevalence [8-10] and burden [11-13] but are often underdetected and undertreated [14,15]. Prevalence estimates of depression (9.3%-23% [16]) and anxiety (2.9%-8.8% [12]) range widely and often vary based on the type and severity of co-occurring chronic conditions. Nevertheless, several studies have suggested that these rates are higher among patients with co-occurring chronic conditions than among those without the conditions [17,18]. Independently, depression has been associated with decrements in physical health outcomes and quality of life [16] for people with chronic conditions, including poor chronic disease self-management [19], worse outcomes for co-occurring physical conditions [20], adverse health system outcomes such as higher use of urgent care [21], and higher costs [22]. Similar impacts have been found for anxiety with chronic conditions [19,23], although this has been less studied. People with chronic conditions face additional challenges in accessing health support in the face of multiple appointments, transportation barriers, and treatment burdens [24]. These barriers may be especially pronounced for those with depression or anxiety, who may have added challenges in accessing support owing to their mental health status and stigma [25].

Digital health interventions (DHIs), that is, health interventions delivered through digital technologies, may help overcome barriers to both delivering mental health care (ie, stigma and access) and care for chronic conditions (ie, communication barriers and lack of appropriate information) [26,27] because of their ability to be accessed remotely, discreetly, and in real time. Existing meta-analytic research [28-34] suggests that DHIs are effective interventions for improving both psychological and disease-specific outcomes for populations with chronic conditions, with small to moderate effect sizes [28,30,32]. The most recent and largest of these meta-analyses reported that self-guided web-based interventions were associated with significant reductions in depression and anxiety compared with usual care or waitlist control, with small effect sizes [34]. However, existing reviews have been limited to a narrow set of therapeutic strategies (eg, self-help [34,35], mindfulness interventions [31,36], cognitive behavioral therapy [CBT]; [28,30,32]), technologies (eg, web-based platforms) [28,30,32,34-36], and outcomes (eg, effectiveness) [28,30-37], leaving the design and implementation of such technologies less clear.

### Objectives

To complement previous reviews, this study sought to examine a wider range of technologies (mobile apps, telemonitoring systems, etc), functions DHIs may serve with respect to depression and anxiety (prevention, detection, or treatment), and study designs (qualitative, quantitative, pilot studies, etc). Specifically, this scoping review aimed to explore what is known about DHIs to prevent, detect, and treat depression or anxiety among people living with chronic conditions. To our knowledge, this is the first scoping review of primary research on DHIs for depression or anxiety in people with chronic conditions.

## *Methods*

### Overview

This scoping review was based on the following framework for conducting scoping reviews as developed by Arksey and O’Malley [38] and refined by Levac et al [39]: (1) identifying the research questions in light of the research purpose; (2) identifying relevant studies while balancing feasibility with comprehensiveness; (3) identifying the study selection criteria using an iterative team approach; (4) charting the data to provide both a numerical summary and thematic analysis; and (5) reporting the results with implications for policy, practice, and research identified [1]. In addition to the framework by Arksey and O’Malley [38] and Levac et al [39], this review adhered to the scoping review guidelines outlined by the JBI [40]. Informed by these frameworks, the actions taken at each stage of the scoping review process are described in the following sections [2].

### Stage 1: Identifying the Research Question

This scoping review was guided by the following research question: What is known about the use of DHIs to support the prevention, detection, or treatment of depression or anxiety among people with chronic conditions? To address this research question, this study sought to (1) describe the nature and extent of DHIs to support the prevention, detection, and treatment of depression or anxiety among people with chronic conditions; (2) describe existing research and its overall findings; and (3) identify gaps and opportunities for future research.

### Stage 2: Identifying Relevant Studies

This scoping review was designed and reported in line with the PRISMA-ScR (Preferred Reporting Items for Systematic

Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [41]. A protocol was developed to guide this review; however, it was not registered. Studies relevant to the research question were identified through searches of the following databases: Embase, CINAHL, PsycINFO, MEDLINE, and the Cochrane Library using terms related to the concepts of “digital health,” “chronic disease,” “depression,” and “anxiety.” As recommended by the Joanna Briggs Institute [40], a 3-step process was used to develop the search strategy. First, initial searches of MEDLINE and PsycINFO were conducted to identify the terms used in the titles and abstracts of the articles. This first step helped identify relevant keywords and subject headings. Second, terms identified in the initial searches were arranged into a search strategy for MEDLINE, which was later tailored and revised for other databases. An iterative team approach was used to develop the search strategy [39] by consulting with a discipline-specific research librarian and seeking feedback from the research team. After tailoring the initial search strategy for MEDLINE to the rest of the databases (Embase, CINAHL, PsycINFO, and the Cochrane Library), searches of all databases were conducted initially in April 2019 and later updated in March 2021 to capture new publications between April 2019 and March 2021. The search strategy for all databases was saved to ensure reproducibility of the search results (see Multimedia Appendix 2 for the search strategy for MEDLINE). In the third step, the reference lists of relevant studies were examined in other studies that could be pertinent to the research question. Journals related to digital health (*Journal of Medical Internet Research* and *Internet Interventions*) were also hand searched for potentially relevant articles.

## Stage 3: Study Selection

The results of the research strategy were saved in the reference management software Mendeley (Elsevier) to identify duplicates and were then exported to Rayyan [42], an internet-based platform designed to expedite the process of screening articles in systematic reviews. As outlined by Arksey and O’Malley [38], studies were selected based on their relevance to the scoping review question rather than their methodological rigor. The studies were reviewed based on the inclusion and exclusion criteria listed in Textbox 1.

**Textbox 1.** Study eligibility criteria.

**Inclusion criteria**

- Article published in English in a peer-reviewed academic journal
- Published after 1990, owing to our interest in newer technologies
- Describe a digital health intervention (DHI) defined as a health intervention delivered via digital technologies including but not limited to web-based platforms, videoconferencing, mobile phone apps, SMS text messages, email, wearable devices, and monitoring sensors
- Study population of individuals aged ≥18 years with one or more of the common chronic conditions identified by the Public Health Agency of Canada (PHAC) [43]: arthritis, asthma, diabetes mellitus, heart disease, chronic obstructive pulmonary disease, cancer, stroke, and Alzheimer disease and related dementias [4]. Although depression and anxiety are also deemed common chronic conditions per PHAC, these conditions were selected because of their high rate of comorbidity with the other common chronic conditions listed [8-10].
- Intervention objective, in whole or in part, related to the prevention, detection, or treatment of depression or anxiety. This was operationalized by studies that either
  - Included a study population with depression or anxiety in addition to one or more of the common chronic conditions outlined by PHAC [43].
  - Explicitly stated that the intervention sought to prevent, detect, or treat depression or anxiety among those with one or more of the common chronic conditions outlined by PHAC [43].

**Exclusion criteria**

- Editorials, case reports, abstracts, posters, or dissertations
- Interventions that solely used the phone calling functionality of telephones [44] (including interactive voice response) were not included in our definition of DHIs stated previously
- Interventions for other mood or mental disorders beyond depression and anxiety
- Studies with a mixed youth and young adult population (eg, [45])
- Lifestyle or survivorship programs

Overall, 2 reviewers (AS and NHS) screened studies yielded by the search strategy by title and abstract and excluded studies if they met any of the exclusion criteria. Rayyan QCRI [42] was used to coordinate screening between the reviewers. Efforts were made to retain studies until full-text review if the reviewers were unsure about the eligibility of the studies with respect to the scoping review criteria. Review articles (ie, systematic reviews, meta-analyses, meta-syntheses, scoping reviews, narrative reviews, rapid reviews, critical reviews, and integrative reviews) were collected, and the reference lists were scanned for potentially relevant articles. Once the reference lists were scanned and potentially relevant articles were identified, the review articles were excluded. The full text of the remaining studies was reviewed according to the inclusion and exclusion criteria. Studies that met the inclusion criteria and those that did not meet the exclusion criteria were included in the scoping review. Studies that met the exclusion criteria were excluded, and the reasons were recorded. Disagreements between the reviewers were resolved through discussion until a consensus was reached, and a third reviewer (ES) was engaged when needed to resolve any remaining conflicts.

### Stage 4: Data Collection

Two reviewers (AS and NHS) independently extracted data from the studies included in the scoping review using a data extraction form. A preliminary data extraction form was drafted in Microsoft Excel with the following column headings: title, year of publication, study details (eg, location, objective, research design, methods, eligibility criteria, target population, conditions, outcomes measured, and findings), and intervention details (eg, name, goal, technology, therapeutic components, and guidance [where any in-person or digital correspondence with a human was considered guided]). The form was modified and revised as necessary during the charting process. Discrepancies between the reviewers regarding the extracted data were resolved through discussion.

### Stage 5: Data Summary and Synthesis of Results

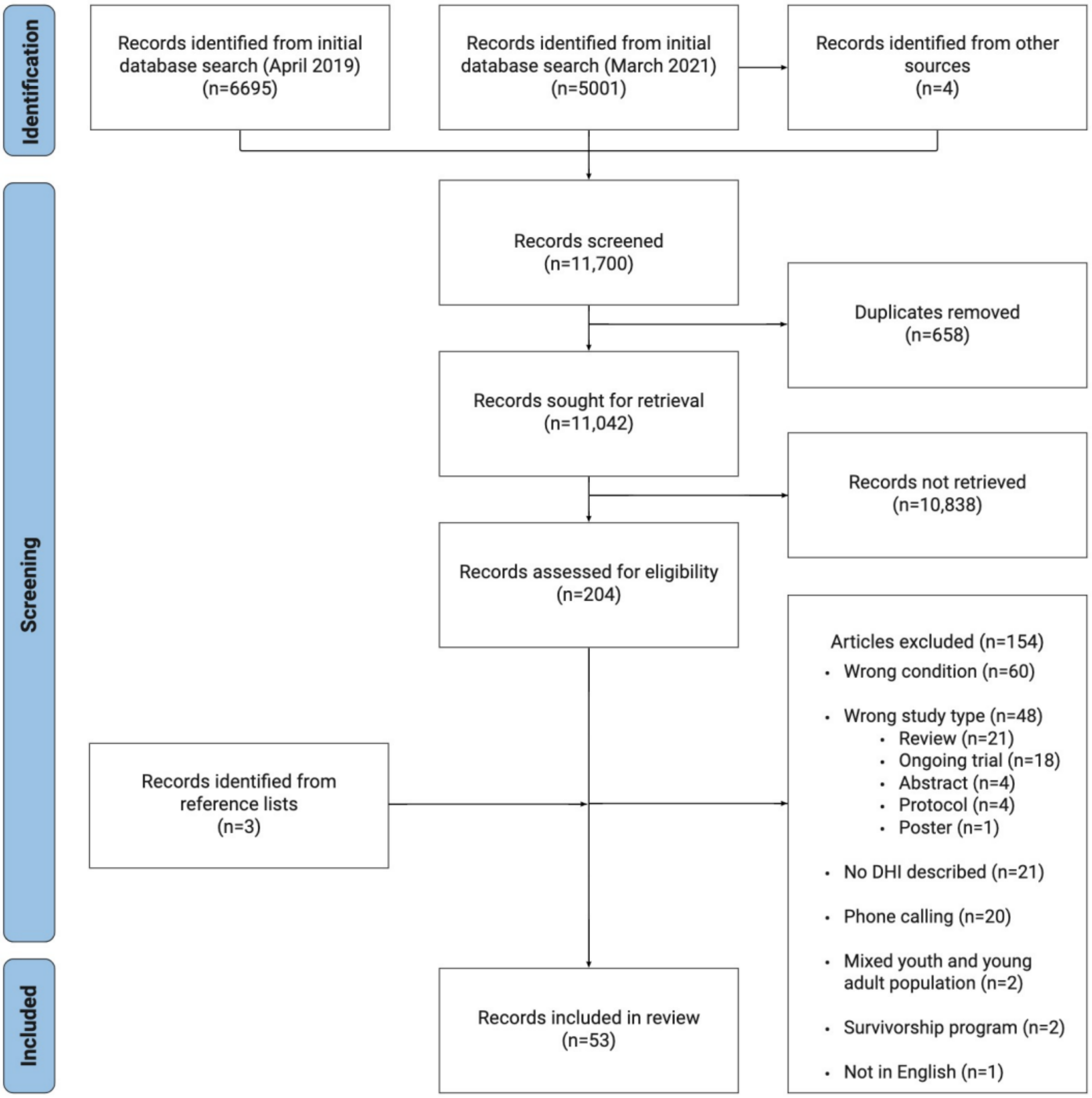
Data analysis was performed using the following steps. First, the results of the nature and extent of the studies and interventions were summarized. Second, details related to the findings of existing studies were collated, and trends were discussed with both reviewers (AS and NHS) to identify opportunities for further research.

## Results

### Overview

Initial searches of 5 databases in April 2019 yielded 6695 results, with one additional article identified through hand searches of a relevant journal (*Journal of Medical Internet Research*; Figure 1). After removing duplicates (n=381), two authors (AS and NHS) screened 6315 records by title and abstract. At this stage, 6228 articles were excluded, including 74 reviews whose reference lists were checked for potentially relevant articles. An additional 3 relevant studies were identified from these reference lists. This left 87 articles to be assessed by full text against the inclusion and exclusion criteria by an author (AS). Of these 87 articles, 61 (70%) were excluded for the following reasons: the study population or intervention goal was not depression or anxiety among individuals with common chronic conditions as identified by the Public Health Agency of Canada [43] (32/61, 52%), the intervention was not a DHI (15/61, 25%), and the intervention relied exclusively on phone calls (14/61, 23%). This resulted in a total of 26 articles. A search update was conducted using the same search strategy in March 2021 for studies published between April 2019 and March 2021. The search update yielded 5001 articles, 4727 (94.52%) and 4610 (92.18%) of which were reviewed in duplicate (AS and NHS) and excluded by title and abstract, respectively. The full text of the remaining 117 studies was reviewed, with 93 (79.5%) studies excluded, resulting in another 24 (20.5%) articles included. An additional 3 articles were identified by screening the reference lists. Overall, 53 articles [46-99] were included in this review (Multimedia Appendix 3 [46,47,50,52,54,55,57-60,62, 64-66,70-72,74,77,78,80,81,84,86,88,90-92,95,96,98,99]). The following sections outline the characteristics of the included studies and report the details related to study objectives 1 and 2.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. DHI: digital health intervention.



## Study Characteristics

### *Publication Frequency*

Table 1 shows the frequency of the included studies by publication year. From 1990 to 2010, no study met our inclusion criteria, highlighting the recent and emerging nature of this area of research. The first study emerged in 2011, with most studies meeting the criteria published in 2019 (16/53, 30%).

**Table 1.** Number of included articles by year of publication (N=53).

| Year | Studies, n (%) |
|---|---|
| 2011 | 3 (6) |
| 2012 | 1 (2) |
| 2013 | 1 (2) |
| 2014 | 1 (2) |
| 2015 | 3 (6) |
| 2016 | 5 (9) |
| 2017 | 4 (8) |
| 2018 | 5 (9) |
| 2019 | 16 (30) |
| 2020 | 13 (25) |
| 2021 | 1 (2) |

### *Setting*

The included studies were predominantly conducted in the United States (15/53, 28%), Sweden (7/53, 13%), Australia (6/53, 11%), Canada (3/53, 6%), and India (4/53, 8%). Other study locations included Germany (3/53, 6%), the Netherlands (2/53, 4%), the United Kingdom (2/53, 4%), China (2/53, 4%), Peru (1/53, 2%), and Jordan (1/53, 2%; Table 2). A total of 4 studies did not specify the study location. On the basis of economy type (as determined by the World Bank classification [100]), studies were largely conducted in high-income countries (38/53, 72%) and to a lesser extent in upper–middle-income countries (4/53, 8%) and lower–middle-income countries (4/53, 8%). There were 3 instances in which studies involved multiple locations; however, these collaborations were between countries with similar economy types: multiple high-income countries (2/53, 4%) and between upper middle-income countries (1/53, 2%).

**Table 2.** Summary of study characteristics (N=53).

| | Studies, n (%) | References |
|---|---|---|
| **Study location** | | |
| United States | 15 (28) | [48,54,58,62,66,69,76,79,82,84,87,89,94,95,97] |
| Sweden | 7 (13) | [47,65,77,80,81,86,92] |
| Australia | 6 (11) | [57,63,85,91,93,96] |
| Canada | 3 (6) | [67,78,83] |
| India | 4 (8) | [61,72,73,75] |
| Germany | 3 (6) | [71,98,99] |
| The Netherlands | 2 (4) | [52,88] |
| United Kingdom | 2 (4) | [49,59] |
| China | 2 (4) | [53,70] |
| Peru | 1 (2) | [90] |
| Jordan | 1 (2) | [51] |
| Location not reported | 4 (8) | [50,55,60,64] |
| **Economy type** | | |
| HIC[a] | 38 (72) | [47-49,52,54,57-59,62,63,65-67,69,71,76-89,91-99] |
| UMIC[b] | 4 (8) | [51,53,68,70] |
| Lower–middle-income country | 4 (8) | [61,72,73,75] |
| Multiple HICs | 2 (4) | [56,74] |
| Multiple UMICs | 1 (2) | [68] |
| **Requirements for study eligibility** | | |
| Access or ownership of a digital device | 13 (25) | [49,59,63,65,73,82,85,87,91,93,95,97,98] |
| Internet access | 9 (17) | [47,52,59,62,63,65,67,91,98] |
| Digital literacy or skills | 6 (11) | [48,57,70,74,82,88] |
| Owning an email address | 3 (6) | [52,74,98] |
| Sample sex distributions: predominantly female (≥60%) | 29 (55) | [49,51,52,54,55,57,60,62-64,68,74,76-81,83,87-91,93,94,97-99] |
| **Research design** | | |
| Randomized controlled trial | 29 (55) | [49,51-53,57-59,62,65,67,70-72,75,78,79,81,83-85,87-89,91,92,95,96,98,99] |
| Quasi-experimental | 5 (9) | [47,60,63,68,76] |
| Grounded theory | 2 (4) | [50,64] |
| Observational | 1 (2) | [73] |
| Case study | 1 (2) | [66] |
| Phenomenological | 1 (2) | [56] |
| **Methods** | | |
| Quantitative | 23 (43) | [52,58,60,62,65,68,73,74,76,78,83,84,88,89,91,93,99,101-106] |
| Qualitative | 9 (17) | [50,55,56,64,66,77,80,90,97] |
| Multi-methods | 8 (15) | [47,48,54,63,69,74,75,89] |

[a] HIC: high-income country.

[b] UMIC: upper–middle-income country.

### *Population*

The sample size ranged from 6 to 3698. Approximately 40% (22/54) of the included studies specified digital requirements for study eligibility, such as access to or ownership of a digital device (13/53, 25%), internet access (9/53, 17%), digital literacy or skills (6/53, 11%), and owning an email address (3/53, 6%).

Most articles reported that the study population was predominantly (≥60%) female (29/53, 55%).

### Methods

Among the studies that reported a research design, the randomized controlled trial design was typically used (29/53, 55%), followed by quasi-experimental (5/53, 9%), grounded theory (2/53, 4%), observational design (1/53, 2%), case study (1/53, 2%), and phenomenological design (1/53, 2%). Different methods were used among the included studies, including quantitative (23/53, 43%), qualitative (9/53, 17%), and multi-methods research (8/53, 15%). A total of 3 articles [61,85,86] described the development of a DHI or the lessons learned in conducting a study but did not state a qualitative methodology.

## Nature and Extent of DHIs

### Intervention Purpose and Digital Technologies

The 53 studies included in this review described 36 unique DHIs. The purpose of the interventions was distributed as follows: prevention (4/53, 8%), detection (5/53, 9%), and treatment (44/53, 83%; Table 3). Over the past decade, the technologies most commonly used to deliver these interventions were web-based platforms (20/53, 38%) and mobile devices (17/53, 32%). Telehealth systems (3/53, 6%), electronic health records (2/53, 4%), and virtual reality devices (1/53, 2%) were also used. Several studies have described DHIs that leverage multiple technologies (10/53, 19%).

**Table 3.** Summary of intervention characteristics (N=53).

| | Studies, n (%) | References |
|---|---|---|
| **Purpose** | | |
| Prevention | 4 (8) | [57,89,93,96] |
| Detection | 5 (8) | [56,67,73,76,82] |
| Treatment | 44 (83) | [47-55,58-66,68-72,74,75,77-81,83-88,90-92,94,95,97-99] |
| **Technology** | | |
| Web based | 20 (38) | [47,49,52,54,58,62,65,74,77,80,81,84-86,88,91-93,96,98] |
| Mobile device | 17 (32) | [48,53,56,61,67-70,73,75,79,82,87,89,90,94,97] |
| Telehealth system | 3 (6) | [66,71,95] |
| Electronic health records | 2 (4) | [72,76] |
| Virtual reality | 1 (2) | [51] |
| Multiple technologies | 10 (19) | [50,55,57,59,60,63,64,78,83,99] |
| **Target conditions (mental)** | | |
| Depression | 36 (68) | [50,52,55-58,60-66,68,69,71-76,78,82-85,87,88,90-93,95,97-99] |
| Anxiety | 3 (6) | [51,59,79] |
| Depression and anxiety | 14 (26) | [47-49,53,54,67,70,77,80,81,86,89,94,96] |
| **Target conditions (physical)** | | |
| Diabetes | 19 (36) | [49,52,61,63,68,72-76,82,85,88,90,91,93,95,98,99] |
| Cancer | 13 (25) | [51,53,54,70,77,79-81,86,87,94,96,97] |
| Hypertension | 5 (9) | [47,61,68,75,90] |
| Heart failure | 3 (6) | [65,71,92] |
| Chronic obstructive pulmonary disease | 2 (4) | [56,67] |
| Stroke | 1 (2) | [59] |
| Any chronic condition | 8 (15) | [50,55,60,62,64,66,69,84] |
| Multiple chronic conditions | 6 (11) | [48,57,58,78,83,89] |
| **Role of digital health intervention** | | |
| Sole intervention | 42 (79) | [47-52,55,57,59,60,62-65,68,69,71,73,74,77-99] |
| Component of intervention | 11 (21) | [53,54,56,58,61,66,67,70,72,75,76] |
| **Intervention components** | | |
| Education | 30 (57) | [46,49,50,52,55,59,60,63,64,66,67,69,70,73,74,77,78,80,81,83-86,88-91,93,94,96] |
| Cognitive behavioral therapy | 21 (40) | [47,49,51,52,59,63,65,69,70,74,75,77-81,83,85,92-94] |
| Behavioral activation | 4 (8) | [46,73,90,91] |
| Problem-solving therapy | 2 (4) | [73,91] |
| Acceptance and commitment therapy | 1 (2) | [54] |
| Monitoring mental health status or symptoms | 14 (26) | [48,49,56,57,63,69,72,78,82,83,85,88,93,98] |
| Peer support | 9 (17) | [50,55,60,64,66,77,80,81,86] |
| Communication with health care providers | 9 (17) | [48,67,71,72,77,80,81,86,95] |
| Mindfulness | 3 (6) | [75,87,97] |
| Chat rooms or forums | 7 (13) | [66,77,80,81,84,86,96] |
| Tailoring: tailored to chronic physical conditions | 19 (36) | [47,48,52,54,59,65,70,73,74,77,79-81,84,86,88,91,92,96] |
| Guidance: guided | 32 (60) | [46,47,50-52,55,58-60,64-66,69-74,77,78,80,81,83,84,86,90-92,94-96,98] |
| **Guidance provider** | | |
| Nurse | 9 (17) | [47,65,68,70,71,86,90,92,96] |

| | Studies, n (%) | References |
|---|---|---|
| Psychologist | 6 (11) | [52,57,74,88,91,98] |
| Certified peer specialist | 4 (8) | [50,55,60,64] |
| Trained lay individual | 2 (4) | [78,83] |
| Allied health professional | 2 (4) | [1,2] |
| Study staff members | 2 (4) | [58,62] |
| Physician | 1 (2) | [59] |
| Multiple professionals | 6 (11) | [54,77,80,81,84,99] |
| Unclear | 1 (2) | [66] |
| **Guidance purpose** | | |
| Responding to questions | 11 (21) | [47,57,62,65,77,80,81,86,92,95,96] |
| Information and feedback | 9 (17) | [52,74,80,81,88,95,96,98,99] |
| Promoting engagement and adherence | 8 (15) | [50,55,57,60,62,64,91,99] |
| Sending reminders | 7 (13) | [47,52,65,74,88,92,98] |
| Offering support | 6 (11) | [50,55,60,64,70,91] |
| Monitoring symptoms | 4 (8) | [72,78,83,95] |
| Training to use the intervention | 4 (8) | [58,68,90,95] |
| Encouragement or positive reinforcement | 4 (8) | [57,68,90,95] |
| Moderating forum | 3 (6) | [80,81,86] |
| Counseling | 2 (4) | [59,72] |
| Check-ins | 1 (2) | [54] |
| Unclear | 2 (4) | [66,84] |
| **Delivery of guidance** | | |
| Combination | 13 (25) | [50,54,55,57,60,64,72,77,90,91,95,98,99] |
| Phone calls | 7 (13) | [59,68,71,78,83,84,96] |
| Emails | 6 (11) | [47,52,65,74,88,92] |
| Web-based messages | 3 (6) | [80,81,86] |
| In-person visits | 2 (4) | [58,62] |
| WeChat messages | 1 (2) | [70] |
| Unclear | 1 (2) | [66] |

### Target Conditions

Most of the studies included in this review sought to address depression (36/53, 68%) among those with chronic conditions, with some studies (14/53, 26%) focusing on both depression and anxiety. Only 6% (3/32) of studies focused exclusively on anxiety. Regarding co-occurring chronic conditions, most interventions were designed for people with diabetes (19/53, 36%) and cancer (13/53, 25%). Other chronic physical conditions represented in the included studies were hypertension (5/53, 9%), heart failure (3/53, 6%), chronic obstructive pulmonary disease (2/53, 4%), and stroke (1/53, 2%). In addition, 8 studies were for any chronic condition, and 6 studies were for multiple chronic conditions (2 or more chronic physical conditions).

### Intervention Design and Components

Of the 53 included studies, 42 (79%) described the DHI as the sole intervention, whereas in 11 (21%) studies, the DHI was a component of the intervention. DHIs commonly included education (30/53, 57%), psychological therapy (28/53, 53%; eg, CBT 21/53, 40%; behavioral activation 4/53, 8%; problem-solving therapy 2/53, 4%; and acceptance and commitment therapy 1/53, 2%), and monitoring of mental health status or symptoms (14/53, 26%). However, peer support (9/53, 17%), communication with health care providers (9/53, 17%), mindfulness (3/53, 6%), and chat rooms or forums (7/53, 13%) were also used. Mental health intervention content was frequently tailored to chronic physical conditions (19/53, 36%; Multimedia Appendix 4 [47-99]).

Most interventions included in this review were guided (32/53, 27%), with the frequency at which guidance was provided ranging from as needed to daily (Multimedia Appendix 5

[47-99]). Among the guided interventions, the provider of guidance varied widely across the studies, including nurse specialists, nurse practitioners, or nurse assistants (9/32, 28%); psychologists (6/32, 19%); certified peer supporters (4/32, 13%); trained lay individuals (2/32, 6%); allied health professionals (2/32, 6%; eg, dieticians or social workers 1/32, 3% and pharmacists 1/32, 3%); study staff members (2/32, 6%); and physicians (1/32, 3%). In 19% (6/32) of studies, guidance was provided by multiple professionals (eg, nurses and psychologists, social workers and psychologists, psychology graduate students, and psychologists). In 3% (1/32) of studies, the provider of guidance was unclear. Guidance served a wide range of functions, including responding to questions (11/32, 34%), providing information and feedback (9/32, 28%), promoting engagement and adherence (8/32, 25%), sending reminders (7/32, 22%), offering support (6/32, 19%), monitoring symptoms (4/32, 13%), training to use the intervention (4/32, 13%), providing encouragement or positive reinforcement (4/32, 13%), moderating a forum (3/32, 9%), counseling (2/32, 6%), and check-ins (1/32, 3%). In some studies (2/32, 6%), the purpose of guidance was unclear. Of the studies describing a guided intervention, most (13/32, 41%) used a combination of modalities (eg, in-person, phone calls, and SMS text messages) to provide guidance. In some studies, guidance was offered exclusively through phone calls (7/32, 22%), emails (6/32, 19%), web-based messages (3/32, 9%), in-person visits (2/32, 6%), and WeChat messages (1/32, 3%). In 3% (1/32) of studies, the method of delivery for guidance was unclear.

### Findings of Existing Studies

The objectives of the included studies ranged from design and development (5/53, 9%), feasibility and acceptability (19/53, 36%), determining effectiveness (17/53, 32%) or efficacy (7/53, 13%) of the DHI, and conducting a secondary analysis to explore predictors, mediators, or moderators of DHI outcomes (6/53, 11%; Table 4). Only 4% (2/32) of studies described the implementation of an intervention, one of which [48] described a planned study but was included because of its discussion of the strategies used to implement the DHI. Other study objectives were to describe the adherence and use of a DHI (1/53, 2%), evaluate a prediction model (1/53, 2%), determine the validity of delivering screening tools via text (1/53, 2%), and analyze SMS text messages from a DHI (1/53, 2%). Several studies reported objectives spanning multiple research stages (8/53, 15%). Among the studies investigating the effectiveness (17/53, 32%) or efficacy (7/53, 13%) of a DHI, 54% (13/32) of studies reported significant changes in mental health outcomes. A study reported significant changes in patients' health behavior (medication adherence [95]) and another in health care provider behavior (depression screening [76]). Multimedia Appendix 3 provides a detailed summary of the findings of each study included in the review.

**Table 4.** Summary of study findings (N=53).

| | Studies, n (%) | References |
|---|---|---|
| **Study objective** | | |
| Design and development | 5 (9) | [47,61,69,74,86] |
| Feasibility and acceptability | 19 (36) | [47,50,54,56,59,60,62-64,66,68,70,76,77,80,85,87,90,97] |
| Effectiveness | 17 (32) | [51-53,58,60,63,65,68,70,72,75,76,78,81,89,93,96] |
| Efficacy | 7 (13) | [49,57,79,84,91,95,99] |
| Secondary analysis | 6 (11) | [71,83,88,92,94,98] |
| Implementation | 2 (4) | [48,70] |
| Other | 3 (6) | [1-3] |
| Impact: significant differences in mental health outcomes | 13 (25) | [51,52,57,58,62,63,68,70,72,78,81,91,99] |

## Discussion

### Principal Findings

This review aimed to explore what is known about DHIs to prevent, detect, and treat depression or anxiety among people with chronic conditions. This review extends the existing meta-analytic evidence of DHIs by reviewing technologies beyond web-based platforms and exploring design and implementation considerations. The findings of this review highlight the significant potential of DHIs to have a profound public health impact on people living with chronic conditions. Per study objective 3, the following section outlines opportunities for further research for this rapidly growing area of investigation: tailoring, guidance, intensity, and stepped care.

### Opportunities for Further Research

#### *Tailoring*

There was mixed evidence among the studies reviewed regarding the value of tailoring the intervention content to chronic physical conditions. The benefits of both condition-specific and generic DHIs were discussed among the studies reviewed. For instance, van Bastelaar et al [74] found that 80% of the patients reported that a diabetes-specific approach for coping with depression intervention was needed. In a previous review, van Beugan et al [30] highlighted the importance of disease-specific tailoring in attributing the larger improvements in disease-related outcomes observed among these interventions for this feature. However, studies in our review also indicated that generic DHIs were beneficial in addressing depression among people with chronic conditions [58,62,91]. It is possible that certain conditions may require a condition-specific approach owing to the existence of condition-specific mental health constructs (eg, diabetes distress) [93]. However, the growing prevalence of multiple chronic conditions globally and emerging literature on transdiagnostic interventions for depression and anxiety [107-109] calls into question the value of disease-specific approaches.

The studies in this review also proposed additional tailoring factors for future consideration. In a qualitative investigation, Igelström et al [80] recommended further tailoring of DHI information, design, and features to the individual with regard to their treatment as well as other factors that appeared to affect the user's experience such as computer experience, internet activity, and interactions with social media. In another study on the same platform [77], patients desired further tailoring to their specific diagnosis, age, sex, treatment, and symptoms, despite the DHI being tailored to the condition (cancer). For the PeerTECH intervention, certified peer specialists personalized SMS text messages on topics such as stories of recovery, medication adherence, coping skills training, and sleep hygiene [50]. Others have raised the possibility of tailoring to participants' preconceptions of the interventions, such as meditation [97] and symptom tracking [80]. Thus, although tailoring can be an important component of DHIs for this population, whether tailoring should be to the *person* instead of the *condition* requires further investigation.

### Guidance

Interventions reported among the included studies were predominately guided; however, both guided and self-guided DHIs were represented among studies reporting effective outcomes, suggesting that outcomes may not be compromised without guidance. The qualifications of individuals providing guidance varied considerably (eg, laypersons to psychiatrists), as did their training and role in the intervention. Similar findings regarding the heterogeneity in the guidance of internet-based CBT for chronic conditions have been reported previously by a meta-analysis by van Began et al [30]. Such variability may arise because of confusion regarding whether guidance constitutes the intervention itself or the supporting context (promoting engagement with technology or study procedures). Indeed, although several interventions relied on guidance to promote engagement with the intervention [50,55,57,60,64,91,99], studies have noted that this limited the scalability of the intervention. Guidance provided by lay individuals and peer supporters was feasible and well accepted [46,60], whereas interventions supported by nurses at times faced challenges in integrating additional workload into their roles [90]. Although several alternatives exist to improve the feasibility of guidance (eg, reduced frequency of guidance, partial guidance, no guidance, blending guided sessions with self-help modules [110], and automated guidance via chatbots [80]), the findings of this review and previous studies underscore the potential for allied health professionals and nonclinicians in guiding DHIs. Nevertheless, clarity is needed regarding the time spent by these individuals and the nature of the support provided (ie, technical support or health-related support) to understand how to support DHIs safely and effectively for this population.

### Intensity

Given the complexity of the population with chronic conditions and comorbid depression or anxiety, identifying for whom DHIs are most appropriate was a significant line of investigation [88,94,98]. Among the studies included in this review, those with more severe depressive symptoms often benefited most from DHI treatment. Puzia et al [94] found that patients with myeloproliferative neoplasms with the poorest baseline global mental health experienced the largest reductions in depression and anxiety symptoms when using a calm mindfulness meditation app. Similarly, despite finding no significant differences in anxiety between a tailored CBT app and control, Greer et al [79] reported that patients with more severe anxiety at baseline benefited the most from the CBT app. Although it is commonly accepted that those with more severe mental health symptoms warrant more intensive treatment [111,112], the findings of this review and previous studies suggest that these individuals also benefit from DHIs [113-115]. Taken together, these findings indicate that individuals with severe symptoms of depression or anxiety can benefit from interventions of varying intensity, especially when intensity is defined by clinician time and contacts. It should be noted that the qualitative findings of the included studies reported that participants found DHIs to be demanding in terms of the time and skills required to read the intervention content and completing exercises in addition to the work involved in learning a new platform and troubleshooting technical and navigation issues [47]. Therefore, an expanded conceptualization of intervention intensity may be fruitful in not only considering intervention intensity in terms of health system resources (eg, clinician time and involvement) but also the patient work involved.

### Stepped Care

Presumably, resolving the mixed findings regarding the previously mentioned factors may not require standardizing the same level of guidance and tailoring for all interventions. It is possible that a variety of DHIs may be delivered and supported through a spectrum of guidance ranging from lay individuals to psychiatrists, with guidance tailored and intensified based on individual needs and preferences. This approach is consistent with a stepped care model that was envisioned to benefit from DHIs in several of the included studies [57,65,77,80,98] and previous reviews on this topic [28,35]. Although over a decade has passed since Cuijpers et al [28] first envisioned the role of DHIs for this population existing within stepped care models, this review identified that the application of these interventions in stepped care models has only recently begun. Nevertheless, the emerging work by Igelström et al [80] demonstrates the promise of this approach in their iCAN-DO stepped care model for patients with cancer, with depression or anxiety symptoms, which included nurse-led (step 1, psychoeducation) and psychologist-led (step 2, internet-based CBT) DHIs. As most patients (60%) in the study did not use the second step of support, a stepped care approach may allow for efficient use of scarce mental health human resources [80]. Early findings on the iCAN-DO stepped care intervention indicated promising results regarding the efficacy of this intervention with reductions in depressive symptoms in some patients with cancer [116]. However, more definitive research is needed to determine whether individuals with single and multiple chronic conditions [57] can benefit from stepped care models, particularly those delivered digitally.

### Limitations

This review had several limitations. First, this review was limited to DHIs for depression or anxiety among people with one or more common chronic conditions identified by the Public Health Agency of Canada [43]. Although not included in this review, other chronic conditions not within the scope of this study (chronic pain, irritable bowel syndrome, tinnitus, epilepsy, etc) have been investigated in previous syntheses [28-32,35,37]. Second, despite engaging a discipline-specific research librarian when developing the search strategy and using multiple search methods (eg, hand searching and reviewing reference lists), the search strategy may have missed relevant studies. Third, because of our interest in DHIs that leverage newer technologies, studies that relied solely on phone calls were excluded. Thus, interventions that may be beneficial, such as telephone-based counseling [117-119] and automated telephone screening [120,121] were not within the scope of this study. As studies with older adults with chronic conditions have identified telephone-based support as a desirable component of mental health support [122], further research is needed on the use of phone calls. Fourth, the restriction to studies that were published in peer-reviewed journals in the English language may have missed publicly available DHIs (eg, mobile apps in app stores) that have not yet been researched or limited the inclusion of studies from non–English-speaking countries. Fifth, to be included, studies in this review must have recruited participants with co-occurring depression or anxiety with chronic conditions or explicitly stated that the goal of the DHI was to prevent, detect, or treat depression or anxiety among people with one or more chronic conditions. Broader aspects of mental health (ie, psychological distress [123], social support [124]), general self-management programs [125], and other mental health conditions (ie, eating disorders and substance use disorders) were beyond the scope of this study and warrant investigation [126].

### Conclusions

Amidst meta-analytic research documenting the potential benefits of DHIs to address depression or anxiety among people living with chronic conditions, this scoping review addresses the paucity of research focusing on the design and implementation considerations of such interventions. This review found that the use of DHIs for depression or anxiety among individuals with chronic conditions is a rapidly growing area of research, with most interventions seeking to provide depression treatment using DHIs that are web based, guided, and tailored to chronic physical conditions. With few studies conducted to date, stepped care models are a promising model to implement efficacious DHIs into standard care, although more definitive research is needed to determine whether individuals with single and multiple chronic conditions can benefit from these models. In constructing such models, questions regarding DHI guidance, tailoring, and intensity are key considerations and require future research. Developments in these areas will aid in realizing the potential of DHIs to transform care for patients with chronic conditions consistent with their holistic health needs.

---

### Conflicts of Interest

None declared.

---

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 25 KB - jmir_v24i9e38030_app1.docx ]

Multimedia Appendix 2
Sample search strategy (MEDLINE).
[DOCX File , 16 KB - jmir_v24i9e38030_app2.docx ]

Multimedia Appendix 3
Summary of digital health interventions.
[DOCX File , 24 KB - jmir_v24i9e38030_app3.docx ]

Multimedia Appendix 4
Summary of digital health intervention components.
[DOCX File , 24 KB - jmir_v24i9e38030_app4.docx ]

Multimedia Appendix 5
Summary of digital health intervention guidance.
[DOCX File , 62 KB - jmir_v24i9e38030_app5.docx ]

---

### References

1. Garin N, Koyanagi A, Chatterji S, Tyrovolas S, Olaya B, Leonardi M, et al. Global multimorbidity patterns: a cross-sectional, population-based, multi-country study. J Gerontol A Biol Sci Med Sci 2016 Feb;71(2):205-214 [FREE Full text] [doi: 10.1093/gerona/glv128] [Medline: 26419978]
2. Afshar S, Roderick PJ, Kowal P, Dimitrov BD, Hill AG. Multimorbidity and the inequalities of global ageing: a cross-sectional study of 28 countries using the World Health Surveys. BMC Public Health 2015 Aug 13;15(1):776 [FREE Full text] [doi: 10.1186/s12889-015-2008-7] [Medline: 26268536]
3. Schenker M, Costa DH. Advances and challenges of health care of the elderly population with chronic diseases in Primary Health Care. Cien Saude Colet 2019 Apr;24(4):1369-1380 [FREE Full text] [doi: 10.1590/1413-81232018244.01222019] [Medline: 31066839]
4. CCDI Steering Committee. The 2017 Canadian chronic disease indicators. Health Promot Chronic Dis Prev Can 2017 Aug;37(8):248-251 [FREE Full text] [doi: 10.24095/hpcdp.37.8.03] [Medline: 28800294]
5. Fortin M, Bravo G, Hudon C, Lapointe L, Almirall J, Dubois M, et al. Relationship between multimorbidity and health-related quality of life of patients in primary care. Qual Life Res 2006 Feb;15(1):83-91. [doi: 10.1007/s11136-005-8661-z] [Medline: 16411033]
6. van Oostrom SH, Picavet HS, de Bruin SR, Stirbu I, Korevaar JC, Schellevis FG, et al. Multimorbidity of chronic diseases and health care utilization in general practice. BMC Fam Pract 2014 Apr 07;15(1):61 [FREE Full text] [doi: 10.1186/1471-2296-15-61] [Medline: 24708798]
7. Bähler C, Huber CA, Brüngger B, Reich O. Multimorbidity, health care utilization and costs in an elderly community-dwelling population: a claims data based observational study. BMC Health Serv Res 2015 Jan 22;15:23 [FREE Full text] [doi: 10.1186/s12913-015-0698-2] [Medline: 25609174]
8. Roy-Byrne PP, Davidson KW, Kessler RC, Asmundson GJ, Goodwin RD, Kubzansky L, et al. Anxiety disorders and comorbid medical illness. Gen Hosp Psychiatry 2008;30(3):208-225. [doi: 10.1016/j.genhosppsych.2007.12.006] [Medline: 18433653]
9. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. Med J Aust 2009 Apr 06;190(S7):S54-S60. [doi: 10.5694/j.1326-5377.2009.tb02471.x] [Medline: 19351294]
10. Gold SM, Köhler-Forsberg O, Moss-Morris R, Mehnert A, Miranda JJ, Bullinger M, et al. Comorbid depression in medical diseases. Nat Rev Dis Primers 2020 Aug 20;6(1):69. [doi: 10.1038/s41572-020-0200-2] [Medline: 32820163]
11. Schmitz N, Wang J, Malla A, Lesage A. Joint effect of depression and chronic conditions on disability: results from a population-based study. Psychosom Med 2007 May;69(4):332-338. [doi: 10.1097/PSY.0b013e31804259e0] [Medline: 17470668]
12. Kessler RC, Ormel J, Demler O, Stang PE. Comorbid mental disorders account for the role impairment of commonly occurring chronic physical disorders: results from the National Comorbidity Survey. J Occup Environ Med 2003 Dec;45(12):1257-1266. [doi: 10.1097/01.jom.0000100000.70011.bb] [Medline: 14665811]
13. Scott KM, Bruffaerts R, Tsang A, Ormel J, Alonso J, Angermeyer MC, et al. Depression-anxiety relationships with chronic physical conditions: results from the World Mental Health Surveys. J Affect Disord 2007 Nov;103(1-3):113-120. [doi: 10.1016/j.jad.2007.01.015] [Medline: 17292480]
14. Calleo J, Stanley MA, Greisinger A, Wehmanen O, Johnson M, Novy D, et al. Generalized anxiety disorder in older medical patients: diagnostic recognition, mental health management and service utilization. J Clin Psychol Med Settings 2009 Jun;16(2):178-185 [FREE Full text] [doi: 10.1007/s10880-008-9144-5] [Medline: 19152056]
15. Hill N, Joubert L, Epstein I. Encouraging self-management in chronically ill patients with co-morbid symptoms of depression and anxiety: an emergency department study and response. Soc Work Health Care 2013;52(2-3):207-221. [doi: 10.1080/00981389.2012.737900] [Medline: 23521385]
16. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet 2007 Sep 08;370(9590):851-858. [doi: 10.1016/S0140-6736(07)61415-9] [Medline: 17826170]
17. Patten SB. Long-term medical conditions and major depression in a Canadian population study at waves 1 and 2. J Affective Disord 2001 Mar;63(1-3):35-41. [doi: 10.1016/s0165-0327(00)00186-5]
18. Gadalla T. Association of comorbid mood disorders and chronic illness with disability and quality of life in Ontario, Canada. Chronic Dis Can 2008;28(4):148-154. [doi: 10.24095/hpcdp.28.4.04]
19. Katon W, Lin EH, Kroenke K. The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. Gen Hosp Psychiatry 2007;29(2):147-155. [doi: 10.1016/j.genhosppsych.2006.11.005] [Medline: 17336664]
20. Stein MB, Cox BJ, Afifi TO, Belik S, Sareen J. Does co-morbid depressive illness magnify the impact of chronic physical illness? A population-based perspective. Psychol Med 2006 May;36(5):587-596. [doi: 10.1017/S0033291706007239] [Medline: 16608557]
21. Dickens C, Katon W, Blakemore A, Khara A, McGowan L, Tomenson B, et al. Does depression predict the use of urgent and unscheduled care by people with long term conditions? A systematic review with meta-analysis. J Psychosom Res 2012 Nov;73(5):334-342. [doi: 10.1016/j.jpsychores.2012.08.018] [Medline: 23062805]

22. Naylor C, Parsonage M, McDaid D, Knapp M, Fossey M, Galea A. Long-term Conditions and Mental Health The Cost of Co-morbidities. London, United Kingdom: The King's Fund; 2012.
23. Härter MC, Conway KP, Merikangas KR. Associations between anxiety disorders and physical illness. Eur Arch Psychiatry Clin Neurosci 2003 Dec;253(6):313-320. [doi: 10.1007/s00406-003-0449-y] [Medline: 14714121]
24. Ho JW, Kuluski K, Im J. "It's a fight to get anything you need" - accessing care in the community from the perspectives of people with multimorbidity. Health Expect 2017 Dec 24;20(6):1311-1319 [FREE Full text] [doi: 10.1111/hex.12571] [Medline: 28544493]
25. DiNapoli EA, Cinna C, Whiteman KL, Fox L, Appelt CJ, Kasckow J. Mental health treatment preferences and challenges of living with multimorbidity from the veteran perspective. Int J Geriatr Psychiatry 2016 Oct 21;31(10):1097-1104 [FREE Full text] [doi: 10.1002/gps.4550] [Medline: 27442187]
26. Mallow JA, Theeke LA, Theeke E, Mallow BK. The effectiveness of mI SMART: a nurse practitioner led technology intervention for multiple chronic conditions in primary care. Int J Nurs Sci 2018 Apr 10;5(2):131-137 [FREE Full text] [doi: 10.1016/j.ijnss.2018.03.009] [Medline: 31406814]
27. Melchiorre MG, Papa R, Rijken M, van Ginneken E, Hujala A, Barbabella F. eHealth in integrated care programs for people with multimorbidity in Europe: insights from the ICARE4EU project. Health Policy 2018 Jan;122(1):53-63 [FREE Full text] [doi: 10.1016/j.healthpol.2017.08.006] [Medline: 28899575]
28. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
29. Matcham F, Rayner L, Hutton J, Monk A, Steel C, Hotopf M. Self-help interventions for symptoms of depression, anxiety and psychological distress in patients with physical illnesses: a systematic review and meta-analysis. Clin Psychol Rev 2014 Mar;34(2):141-157 [FREE Full text] [doi: 10.1016/j.cpr.2014.01.005] [Medline: 24508685]
30. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp H, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014 Mar 27;16(3):e88 [FREE Full text] [doi: 10.2196/jmir.2777] [Medline: 24675372]
31. Mikolasek M, Berg J, Witt CM, Barth J. Effectiveness of mindfulness- and relaxation-based eHealth interventions for patients with medical conditions: a systematic review and synthesis. Int J Behav Med 2018 Feb;25(1):1-16. [doi: 10.1007/s12529-017-9679-7] [Medline: 28752414]
32. Mehta S, Peynenburg VA, Hadjistavropoulos HD. Internet-delivered cognitive behaviour therapy for chronic health conditions: a systematic review and meta-analysis. J Behav Med 2019 Apr;42(2):169-187. [doi: 10.1007/s10865-018-9984-x] [Medline: 30387008]
33. Sasseville M, LeBlanc A, Tchuente J, Boucher M, Dugas M, Gisèle M, et al. The impact of technology systems and professional support in digital mental health interventions: a secondary meta-analysis. medRxiv 2021 Apr [FREE Full text] [doi: 10.1101/2021.04.12.21255333]
34. White V, Linardon J, Stone JE, Holmes-Truscott E, Olive L, Mikocka-Walus A, et al. Online psychological interventions to reduce symptoms of depression, anxiety, and general distress in those with chronic health conditions: a systematic review and meta-analysis of randomized controlled trials. Psychol Med 2020 Jul 17;52(3):548-573. [doi: 10.1017/s0033291720002251]
35. Beatty L, Lambert S. A systematic review of internet-based self-help therapeutic interventions to improve distress and disease-control among adults with chronic health conditions. Clin Psychol Rev 2013 Jun;33(4):609-622. [doi: 10.1016/j.cpr.2013.03.004] [Medline: 23603521]
36. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303 [FREE Full text] [doi: 10.2196/jmir.7487] [Medline: 28860106]
37. Sasseville M, LeBlanc A, Boucher M, Dugas M, Mbemba G, Tchuente J, et al. Digital health interventions for the management of mental health in people with chronic diseases: a rapid review. BMJ Open 2021 Apr 05;11(4):e044437 [FREE Full text] [doi: 10.1136/bmjopen-2020-044437] [Medline: 33820786]
38. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Social Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
39. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(1):69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
40. The Joanna Briggs Institute. Joanna Briggs Institute Reviewers' Manual: 2014 edition / Supplement. Adelaide, Australia: The Joanna Briggs Institute; 2014.
41. 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]
42. 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]

43. Varin M, Baker M, Palladino E, Lary T. Canadian Chronic Disease Indicators, 2019 - Updating the data and taking into account mental health. Health Promot Chronic Dis Prev Can 2019 Oct;39(10):281-288 [FREE Full text] [doi: 10.24095/hpcdp.39.10.02] [Medline: 31600041]
44. Ito H, Sato T, Satoh-Asahara N, Noda M. Impact of medication adherence on renal function in comorbid patients with type 2 diabetes and depression: protocol for a cohort study. BMC Fam Pract 2015 Sep 15;16:124 [FREE Full text] [doi: 10.1186/s12875-015-0339-1] [Medline: 26374104]
45. Lunkenheimer F, Domhardt M, Geirhos A, Kilian R, Mueller-Stierlin AS, Holl RW, COACH consortium. Effectiveness and cost-effectiveness of guided internet- and mobile-based CBT for adolescents and young adults with chronic somatic conditions and comorbid depression and anxiety symptoms (youthCOACH): study protocol for a multicentre randomized controlled trial. Trials 2020 Mar 12;21(1):253 [FREE Full text] [doi: 10.1186/s13063-019-4041-9] [Medline: 32164723]
46. McCusker J, Yaffe M, Faria R, Lambert S, Li M, Poirier-Bisson J, et al. Phase II trial of a depression self-care intervention for adult cancer survivors. Eur J Cancer Care (Engl) 2018 Jan 06;27(1):e12763. [doi: 10.1111/ecc.12763] [Medline: 28984000]
47. Lundgren J, Andersson G, Dahlström Ö, Jaarsma T, Köhler AK, Johansson P. Internet-based cognitive behavior therapy for patients with heart failure and depressive symptoms: a proof of concept study. Patient Educ Couns 2015 Aug;98(8):935-942. [doi: 10.1016/j.pec.2015.04.013] [Medline: 25990216]
48. Kearney SM, Williams K, Nikolajski C, Park MJ, Kraemer KL, Landsittel D, et al. Stakeholder impact on the implementation of integrated care: opportunities to consider for patient-centered outcomes research. Contemp Clin Trials 2021 Feb;101:106256. [doi: 10.1016/j.cct.2020.106256] [Medline: 33383229]
49. Baldwin PA, Sanatkar S, Clarke J, Fletcher S, Gunn J, Wilhelm K, et al. A web-based mental health intervention to improve social and occupational functioning in adults with type 2 diabetes (the springboard trial): 12-month outcomes of a randomized controlled trial. J Med Internet Res 2020 Dec 01;22(12):e16729 [FREE Full text] [doi: 10.2196/16729] [Medline: 33258790]
50. Fortuna KL, Storm M, Aschbrenner KA, Bartels SJ. Integration of peer philosophy into a standardized self-management mobile health intervention. Psychiatr Q 2018 Dec;89(4):795-800 [FREE Full text] [doi: 10.1007/s11126-018-9578-3] [Medline: 29671155]
51. Bani Mohammad E, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: a randomized control trial. Pall Supp Care 2018 Sep 10;17(1):29-34. [doi: 10.1017/s1478951518000639]
52. van Bastelaar KM, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2011 Feb;34(2):320-325 [FREE Full text] [doi: 10.2337/dc10-1248] [Medline: 21216855]
53. Ye Z, Chen J, Zhang Y, Hu X, Xuan Z, Yang S, et al. Video education reduces pain and anxiety levels in cancer patients who first use fentanyl transdermal patch: a randomized controlled trial. Drug Des Devel Ther 2020;14:3477-3483 [FREE Full text] [doi: 10.2147/DDDT.S264112] [Medline: 32921984]
54. Arch JJ, Fishbein JN, Ferris MC, Mitchell JL, Levin ME, Slivjak ET, et al. Acceptability, feasibility, and efficacy potential of a multimodal acceptance and commitment therapy intervention to address psychosocial and advance care planning needs among anxious and depressed adults with metastatic cancer. J Palliat Med 2020 Oct 01;23(10):1380-1385. [doi: 10.1089/jpm.2019.0398] [Medline: 31905307]
55. Fortuna KL, Naslund JA, Aschbrenner KA, Lohman MC, Storm M, Batsis JA, et al. Text message exchanges between older adults with serious mental illness and older certified peer specialists in a smartphone-supported self-management intervention. Psychiatr Rehabil J 2019 Mar;42(1):57-63 [FREE Full text] [doi: 10.1037/prj0000305] [Medline: 30010355]
56. Kayyali R, Savickas V, Spruit MA, Kaimakamis E, Siva R, Costello RW, et al. Qualitative investigation into a wearable system for chronic obstructive pulmonary disease: the stakeholders' perspective. BMJ Open 2016 Aug 31;6(8):e011657 [FREE Full text] [doi: 10.1136/bmjopen-2016-011657] [Medline: 27580831]
57. Read J, Sharpe L, Burton AL, Arean PA, Raue PJ, McDonald S, et al. A randomized controlled trial of internet-delivered cognitive behaviour therapy to prevent the development of depressive disorders in older adults with multimorbidity. J Affect Disord 2020 Mar 01;264:464-473. [doi: 10.1016/j.jad.2019.11.077] [Medline: 31767215]
58. Druss BG, Ji X, Glick G, von Esenwein SA. Randomized trial of an electronic personal health record for patients with serious mental illnesses. Am J Psychiatry 2014 Mar;171(3):360-368. [doi: 10.1176/appi.ajp.2013.13070913] [Medline: 24435025]
59. Chun HY, Carson AJ, Tsanas A, Dennis MS, Mead GE, Calabria C, et al. Telemedicine cognitive behavioral therapy for anxiety after stroke: proof-of-concept randomized controlled trial. Stroke 2020 Aug;51(8):2297-2306 [FREE Full text] [doi: 10.1161/STROKEAHA.120.029042] [Medline: 32576090]
60. Fortuna KL, DiMilia PR, Lohman MC, Bruce ML, Zubritsky CD, Halaby MR, et al. Feasibility, acceptability, and preliminary effectiveness of a peer-delivered and technology supported self-management intervention for older adults with serious mental illness. Psychiatr Q 2018 Jun;89(2):293-305 [FREE Full text] [doi: 10.1007/s11126-017-9534-7] [Medline: 28948424]
61. Jindal D, Gupta P, Jha D, Ajay VS, Goenka S, Jacob P, et al. Development of mWellcare: an mHealth intervention for integrated management of hypertension and diabetes in low-resource settings. Glob Health Action 2018;11(1):1517930 [FREE Full text] [doi: 10.1080/16549716.2018.1517930] [Medline: 30253691]

62. Wilson M, Hewes C, Barbosa-Leiker C, Mason A, Wuestney KA, Shuen JA, et al. Engaging adults with chronic disease in online depressive symptom self-management. West J Nurs Res 2018 Jun;40(6):834-853. [doi: 10.1177/0193945916689068] [Medline: 28322662]
63. Clarke J, Proudfoot J, Ma H. Mobile phone and web-based cognitive behavior therapy for depressive symptoms and mental health comorbidities in people living with diabetes: results of a feasibility study. JMIR Ment Health 2016 May 31;3(2):e23 [FREE Full text] [doi: 10.2196/mental.5131] [Medline: 27245948]
64. Fortuna KL, Storm M, Naslund JA, Chow P, Aschbrenner KA, Lohman MC, et al. Certified peer specialists and older adults with serious mental illness' perspectives of the impact of a peer-delivered and technology-supported self-management intervention. J Nerv Ment Dis 2018 Nov;206(11):875-881 [FREE Full text] [doi: 10.1097/NMD.0000000000000896] [Medline: 30371642]
65. Lundgren JG, Dahlström Ö, Andersson G, Jaarsma T, Kärner Köhler A, Johansson P. The effect of guided web-based cognitive behavioral therapy on patients with depressive symptoms and heart failure: a pilot randomized controlled trial. J Med Internet Res 2016 Aug 03;18(8):e194 [FREE Full text] [doi: 10.2196/jmir.5556] [Medline: 27489077]
66. Sorocco KH, Bratkovich KL, Wingo R, Qureshi SM, Mason PJ. Integrating care coordination home telehealth and home based primary care in rural Oklahoma: a pilot study. Psychol Serv 2013 Aug;10(3):350-352. [doi: 10.1037/a0032785] [Medline: 23937085]
67. Whelan ME, Velardo C, Rutter H, Tarassenko L, Farmer AJ. Mood monitoring over one year for people with chronic obstructive pulmonary disease using a mobile health system: retrospective analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2019 Nov 22;7(11):e14946 [FREE Full text] [doi: 10.2196/14946] [Medline: 31755872]
68. Menezes P, Quayle J, Garcia Claro H, da Silva S, Brandt LR, Diez-Canseco F, et al. Use of a mobile phone app to treat depression comorbid with hypertension or diabetes: a pilot study in Brazil and Peru. JMIR Ment Health 2019 Apr 26;6(4):e11698 [FREE Full text] [doi: 10.2196/11698] [Medline: 31025949]
69. Fortuna KL, Lohman MC, Gill LE, Bruce ML, Bartels SJ. Adapting a psychosocial intervention for smartphone delivery to middle-aged and older adults with serious mental illness. Am J Geriatr Psychiatry 2017 Aug;25(8):819-828 [FREE Full text] [doi: 10.1016/j.jagp.2016.12.007] [Medline: 28169129]
70. Sui Y, Wang T, Wang X. The impact of WeChat app-based education and rehabilitation program on anxiety, depression, quality of life, loss of follow-up and survival in non-small cell lung cancer patients who underwent surgical resection. Eur J Oncol Nurs 2020 Apr;45:101707. [doi: 10.1016/j.ejon.2019.101707] [Medline: 32078926]
71. Koehler J, Stengel A, Hofmann T, Wegscheider K, Koehler K, Sehner S, et al. Telemonitoring in patients with chronic heart failure and moderate depressed symptoms: results of the Telemedical Interventional Monitoring in Heart Failure (TIM-HF) study. Eur J Heart Fail 2021 Jan;23(1):186-194 [FREE Full text] [doi: 10.1002/ejhf.2025] [Medline: 33063412]
72. Ali MK, Chwastiak L, Poongothai S, Emmert-Fees KM, Patel SA, Anjana RM, INDEPENDENT Study Group. Effect of a collaborative care model on depressive symptoms and glycated hemoglobin, blood pressure, and serum cholesterol among patients with depression and diabetes in India: the INDEPENDENT randomized clinical trial. JAMA 2020 Aug 18;324(7):651-662 [FREE Full text] [doi: 10.1001/jama.2020.11747] [Medline: 32809002]
73. Sarda A, Munuswamy S, Sarda S, Subramanian V. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: cross-sectional observational study. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e11041 [FREE Full text] [doi: 10.2196/11041] [Medline: 30694197]
74. van Bastelaar K, Cuijpers P, Pouwer F, Riper H, Snoek FJ. Development and reach of a web-based cognitive behavioural therapy programme to reduce symptoms of depression and diabetes-specific distress. Patient Educ Couns 2011 Jul;84(1):49-55. [doi: 10.1016/j.pec.2010.06.013] [Medline: 20619577]
75. Prabhakaran D, Jha D, Prieto-Merino D, Roy A, Singh K, Ajay VS, Members of the Research Steering Committee,Investigators,Members of the Data SafetyMonitoring Board. Effectiveness of an mHealth-based electronic decision support system for integrated management of chronic conditions in primary care: the mWellcare cluster-randomized controlled trial. Circulation 2019 Jan 15;139(3):380-391 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.118.038192] [Medline: 30586732]
76. Bajracharya P, Summers L, Amatya AK, DeBlieck C. Implementation of a depression screening protocol and tools to improve screening for depression in patients with diabetes in the primary care setting. J Nurse Pract 2016 Nov;12(10):690-696. [doi: 10.1016/j.nurpra.2016.08.009]
77. Hauffman A, Alfonsson S, Igelström H, Johansson B. Experiences of internet-based stepped care in individuals with cancer and concurrent symptoms of anxiety and depression: qualitative exploration conducted alongside the U-CARE adultCan randomized controlled trial. J Med Internet Res 2020 Mar 30;22(3):e16547 [FREE Full text] [doi: 10.2196/16547] [Medline: 32224483]
78. McCusker J, Cole MG, Yaffe M, Strumpf E, Sewitch M, Sussman T, et al. A randomized trial of a depression self-care toolkit with or without lay telephone coaching for primary care patients with chronic physical conditions. Gen Hosp Psychiatry 2016;40:75-83. [doi: 10.1016/j.genhosppsych.2016.02.004] [Medline: 27067224]
79. Greer JA, Jacobs J, Pensak N, MacDonald JJ, Fuh C, Perez GK, et al. Randomized trial of a tailored cognitive-behavioral therapy mobile application for anxiety in patients with incurable cancer. Oncologist 2019 Aug;24(8):1111-1120 [FREE Full text] [doi: 10.1634/theoncologist.2018-0536] [Medline: 30683710]

80. Igelström H, Hauffman A, Alfonsson S, Sjöström J, Cajander Å, Johansson B. User experiences of an internet-based stepped-care intervention for individuals with cancer and concurrent symptoms of anxiety or depression (the U-CARE adultCan trial): qualitative study. J Med Internet Res 2020 May 19;22(5):e16604 [FREE Full text] [doi: 10.2196/16604] [Medline: 32427108]
81. Hauffman A, Alfonsson S, Bill-Axelson A, Bergkvist L, Forslund M, Mattsson S, et al. Cocreated internet-based stepped care for individuals with cancer and concurrent symptoms of anxiety and depression: results from the U-CARE AdultCan randomized controlled trial. Psychooncology 2020 Dec 22;29(12):2012-2018 [FREE Full text] [doi: 10.1002/pon.5489] [Medline: 32691455]
82. Jin H, Wu S. Text messaging as a screening tool for depression and related conditions in underserved, predominantly minority safety net primary care patients: validity study. J Med Internet Res 2020 Mar 26;22(3):e17282 [FREE Full text] [doi: 10.2196/17282] [Medline: 32213473]
83. McCusker J, Lambert S, Cole M, Ciampi A, Strumpf E, Freeman EE, et al. Activation and self-efficacy in a randomized trial of a depression self-care intervention. Health Educ Behav 2016 Dec;43(6):716-725 [FREE Full text] [doi: 10.1177/1090198116637601] [Medline: 27179288]
84. Steel J, Geller DA, Tsung A, Marsh JW, Dew MA, Spring M, et al. Randomized controlled trial of a collaborative care intervention to manage cancer-related symptoms: lessons learned. Clin Trials 2011 Jun;8(3):298-310 [FREE Full text] [doi: 10.1177/1740774511402368] [Medline: 21730078]
85. Fletcher S, Clarke J, Sanatkar S, Baldwin P, Gunn J, Zwar N, et al. Recruiting to a randomized controlled trial of a web-based program for people with type 2 diabetes and depression: lessons learned at the intersection of e-mental health and primary care. J Med Internet Res 2019 May 24;21(5):e12793 [FREE Full text] [doi: 10.2196/12793] [Medline: 31127718]
86. Hauffman A, Alfonsson S, Mattsson S, Forslund M, Bill-Axelson A, Nygren P, et al. The development of a nurse-led internet-based learning and self-care program for cancer patients with symptoms of anxiety and depression—a part of U-CARE. Cancer Nurs 2017;40(5):E9-16. [doi: 10.1097/ncc.0000000000000402]
87. Huberty J, Eckert R, Larkey L, Kurka J, Rodríguez De Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. JMIR Form Res 2019 Apr 29;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]
88. van Bastelaar KM, Pouwer F, Cuijpers P, Riper H, Twisk JW, Snoek FJ. Is a severe clinical profile an effect modifier in a web-based depression treatment for adults with type 1 or type 2 diabetes? Secondary analyses from a randomized controlled trial. J Med Internet Res 2012 Jan 05;14(1):e2 [FREE Full text] [doi: 10.2196/jmir.1657] [Medline: 22262728]
89. Wittink MN, Walsh P, Yilmaz S, Mendoza M, Street RL, Chapman BP, et al. Patient priorities and the doorknob phenomenon in primary care: can technology improve disclosure of patient stressors? Patient Educ Couns 2018 Feb;101(2):214-220 [FREE Full text] [doi: 10.1016/j.pec.2017.08.004] [Medline: 28844522]
90. Brandt LR, Hidalgo L, Diez-Canseco F, Araya R, Mohr DC, Menezes PR, et al. Addressing depression comorbid with diabetes or hypertension in resource-poor settings: a qualitative study about user perception of a nurse-supported smartphone app in Peru. JMIR Ment Health 2019 Jun 18;6(6):e11701 [FREE Full text] [doi: 10.2196/11701] [Medline: 31215511]
91. Newby J, Robins L, Wilhelm K, Smith J, Fletcher T, Gillis I, et al. Web-based cognitive behavior therapy for depression in people with diabetes mellitus: a randomized controlled trial. J Med Internet Res 2017 May 15;19(5):e157 [FREE Full text] [doi: 10.2196/jmir.7274] [Medline: 28506956]
92. Johansson P, Jaarsma T, Andersson G, Lundgren J. The impact of internet-based cognitive behavioral therapy and depressive symptoms on self-care behavior in patients with heart failure: a secondary analysis of a randomised controlled trial. Int J Nurs Stud 2021 Apr;116:103454. [doi: 10.1016/j.ijnurstu.2019.103454] [Medline: 31727306]
93. Clarke J, Sanatkar S, Baldwin PA, Fletcher S, Gunn J, Wilhelm K, et al. A web-based cognitive behavior therapy intervention to improve social and occupational functioning in adults with type 2 diabetes (The SpringboarD Trial): randomized controlled trial. J Med Internet Res 2019 May 21;21(5):e12246 [FREE Full text] [doi: 10.2196/12246] [Medline: 31115345]
94. Puzia ME, Huberty J, Eckert R, Larkey L, Mesa R. Associations between global mental health and response to an app-based meditation intervention in myeloproliferative neoplasm patients. Integr Cancer Ther 2020;19:1534735420927780 [FREE Full text] [doi: 10.1177/1534735420927780] [Medline: 32564631]
95. Cohen LB, Taveira TH, Wu W, Pirraglia PA. Pharmacist-led telehealth disease management program for patients with diabetes and depression. J Telemed Telecare 2020 Jun;26(5):294-302. [doi: 10.1177/1357633X18822575] [Medline: 30691328]
96. Stevenson W, Bryant J, Watson R, Sanson-Fisher R, Oldmeadow C, Henskens F, et al. A multi-center randomized controlled trial to reduce unmet needs, depression, and anxiety among hematological cancer patients and their support persons. J Psychosoc Oncol 2020 Dec 13;38(3):272-292. [doi: 10.1080/07347332.2019.1692991] [Medline: 31833452]
97. Huberty J, Eckert R, Larkey L, Joeman L, Mesa R. Experiences of using a consumer-based mobile meditation app to improve fatigue in myeloproliferative patients: qualitative study. JMIR Cancer 2019 Jul 22;5(2):e14292 [FREE Full text] [doi: 10.2196/14292] [Medline: 31333197]
98. Schlicker S, Weisel KK, Buntrock C, Berking M, Nobis S, Lehr D, et al. Do nonsuicidal severely depressed individuals with diabetes profit from internet-based guided self-help? Secondary analyses of a pragmatic randomized trial. J Diabetes Res 2019;2019:2634094 [FREE Full text] [doi: 10.1155/2019/2634094] [Medline: 31218230]

99. Nobis S, Lehr D, Ebert DD, Baumeister H, Snoek F, Riper H, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. Diabetes Care 2015 May;38(5):776-783. [doi: 10.2337/dc14-1728] [Medline: 25710923]
100. How are the income group thresholds determined? World Bank Data. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/378833-how-are-the-income-group-thresholds-determined [accessed 2019-10-28]
101. Proudfoot J, Clarke J, Gunn J, Fletcher S, Sanatkar S, Wilhelm K, et al. A web-based public health intervention to reduce functional impairment and depressive symptoms in adults with type 2 diabetes (the Springboard Trial): randomized controlled trial protocol. JMIR Res Protoc 2017 Aug 03;6(8):e145 [FREE Full text] [doi: 10.2196/resprot.7348] [Medline: 28778848]
102. Cockayne NL, Glozier N, Naismith SL, Christensen H, Neal B, Hickie IB. Internet-based treatment for older adults with depression and co-morbid cardiovascular disease: protocol for a randomised, double-blind, placebo controlled trial. BMC Psychiatry 2011 Jan 14;11:10 [FREE Full text] [doi: 10.1186/1471-244X-11-10] [Medline: 21232162]
103. Jin H, Wu S. Screening depression and related conditions via text messaging versus interview assessment: protocol for a randomized study. JMIR Res Protoc 2019 Mar 29;8(3):e12392 [FREE Full text] [doi: 10.2196/12392] [Medline: 30924787]
104. Monreal-Bartolomé A, Barceló-Soler A, Castro A, Pérez-Ara MA, Gili M, Mayoral F, et al. Efficacy of a blended low-intensity internet-delivered psychological programme in patients with multimorbidity in primary care: study protocol for a randomized controlled trial. BMC Psychiatry 2019 Feb 11;19(1):66 [FREE Full text] [doi: 10.1186/s12888-019-2037-3] [Medline: 30744610]
105. Robins L, Newby J, Wilhelm K, Smith J, Fletcher T, Ma T, et al. Internet-delivered cognitive behaviour therapy for depression in people with diabetes: study protocol for a randomised controlled trial. BMJ Open Diabetes Res Care 2015;3(1):e000144 [FREE Full text] [doi: 10.1136/bmjdrc-2015-000144] [Medline: 26688735]
106. van Bastelaar KM, Pouwer F, Cuijpers P, Twisk JW, Snoek FJ. Web-based cognitive behavioural therapy (W-CBT) for diabetes patients with co-morbid depression: design of a randomised controlled trial. BMC Psychiatry 2008 Feb 19;8:9 [FREE Full text] [doi: 10.1186/1471-244X-8-9] [Medline: 18284670]
107. Dindo L, Van Liew JR, Arch JJ. Acceptance and commitment therapy: a transdiagnostic behavioral intervention for mental health and medical conditions. Neurotherapeutics 2017 Jul;14(3):546-553 [FREE Full text] [doi: 10.1007/s13311-017-0521-3] [Medline: 28271287]
108. Newby JM, Twomey C, Yuan Li SS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2016 Jul 15;199:30-41. [doi: 10.1016/j.jad.2016.03.018] [Medline: 27060430]
109. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
110. Shapira S, Yeshua-Katz D, Cohn-Schwartz E, Aharonson-Daniel L, Sarid O, Clarfield AM. A pilot randomized controlled trial of a group intervention via Zoom to relieve loneliness and depressive symptoms among older persons during the COVID-19 outbreak. Internet Interv 2021 Apr;24:100368 [FREE Full text] [doi: 10.1016/j.invent.2021.100368] [Medline: 33527072]
111. Webb CA, Rosso IM, Rauch SL. Internet-based cognitive-behavioral therapy for depression: current progress and future directions. Harv Rev Psychiatry 2017;25(3):114-122 [FREE Full text] [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
112. Weisel KK, Lehr D, Heber E, Zarski A, Berking M, Riper H, et al. Severely burdened individuals do not need to be excluded from internet-based and mobile-based stress management: effect modifiers of treatment outcomes from three randomized controlled trials. J Med Internet Res 2018 Jun 19;20(6):e211 [FREE Full text] [doi: 10.2196/jmir.9387] [Medline: 29921562]
113. Button KS, Wiles NJ, Lewis G, Peters TJ, Kessler D. Factors associated with differential response to online cognitive behavioural therapy. Soc Psychiatry Psychiatr Epidemiol 2012 May 4;47(5):827-833. [doi: 10.1007/s00127-011-0389-1] [Medline: 21541696]
114. Spek V, Nyklícek I, Cuijpers P, Pop V. Predictors of outcome of group and internet-based cognitive behavior therapy. J Affect Disord 2008 Jan;105(1-3):137-145. [doi: 10.1016/j.jad.2007.05.001] [Medline: 17543392]
115. Warmerdam L, Van Straten A, Twisk J, Cuijpers P. Predicting outcome of internet-based treatment for depressive symptoms. Psychother Res 2013;23(5):559-567. [doi: 10.1080/10503307.2013.807377] [Medline: 23848944]
116. Johansson B, Hauffman A, Bill-Axelsson A, Alfonsson S, von Essen L, Nygren H, et al. Internet-based stepped care for individuals with cancer and concurrent anxiety or depression symptoms: results from a randomized controlled trial. Ann Oncol 2019 Oct;30:v667-v668. [doi: 10.1093/annonc/mdz262.002]
117. Cassin SE, Sockalingam S, Du C, Wnuk S, Hawa R, Parikh SV. A pilot randomized controlled trial of telephone-based cognitive behavioural therapy for preoperative bariatric surgery patients. Behav Res Ther 2016 May;80:17-22 [FREE Full text] [doi: 10.1016/j.brat.2016.03.001] [Medline: 26990279]
118. Sockalingam S, Cassin SE, Wnuk S, Du C, Jackson T, Hawa R, et al. A pilot study on telephone cognitive behavioral therapy for patients six-months post-bariatric surgery. Obes Surg 2017 Mar;27(3):670-675 [FREE Full text] [doi: 10.1007/s11695-016-2322-x] [Medline: 27491293]

119. Costa-Dookhan KA, Leung SE, Cassin SE, Sockalingam S. Psychosocial predictors of response to telephone-based cognitive behavioural therapy in bariatric surgery patients. Can J Diabetes 2020 Apr;44(3):236-240. [doi: 10.1016/j.jcjd.2019.06.008] [Medline: 31447318]
120. Ramirez M, Wu S, Jin H, Ell K, Gross-Schulman S, Myerchin Sklaroff L, et al. Automated remote monitoring of depression: acceptance among low-income patients in diabetes disease management. JMIR Ment Health 2016 Jan 25;3(1):e6 [FREE Full text] [doi: 10.2196/mental.4823] [Medline: 26810139]
121. Wu S, Ell K, Gross-Schulman SG, Sklaroff LM, Katon WJ, Nezu AM, et al. Technology-facilitated depression care management among predominantly Latino diabetes patients within a public safety net care system: comparative effectiveness trial design. Contemp Clin Trials 2014 Mar;37(2):342-354. [doi: 10.1016/j.cct.2013.11.002] [Medline: 24215775]
122. Grundberg A, Ebbeskog B, Gustafsson SA, Religa D. Mental health-promoting dialogues from the perspective of community-dwelling seniors with multimorbidity. J Multidiscip Healthc 2014;7:189-199 [FREE Full text] [doi: 10.2147/JMDH.S59307] [Medline: 24812516]
123. Bisseling E, Cillessen L, Spinhoven P, Schellekens M, Compen F, van der Lee M, et al. Development of the therapeutic alliance and its association with internet-based mindfulness-based cognitive therapy for distressed cancer patients: secondary analysis of a multicenter randomized controlled trial. J Med Internet Res 2019 Oct 18;21(10):e14065 [FREE Full text] [doi: 10.2196/14065] [Medline: 31628791]
124. Lally RM, Kupzyk K, Gallo S, Berry D. Use of an unguided, web-based distress self-management program after breast cancer diagnosis: sub-analysis of CaringGuidance pilot study. J Med Internet Res 2020 Jul 06;22(7):e19734 [FREE Full text] [doi: 10.2196/19734] [Medline: 32628117]
125. Zhang X, Tan SS, Fierloos I, Zanutto O, Alhambra-Borrás T, Vasiljev V, et al. Evaluation design of the social engagement framework for addressing the chronic-disease-challenge (SEFAC): a mindfulness-based intervention to promote the self-management of chronic conditions and a healthy lifestyle. BMC Public Health 2019 May 30;19(1):664 [FREE Full text] [doi: 10.1186/s12889-019-6979-7] [Medline: 31146709]
126. Murray E, Kerr C, Stevenson F, Gore C, Nazareth I. Internet interventions can meet the emotional needs of patients and carers managing long-term conditions. J Telemed Telecare 2016 Dec 02;13(1_suppl):42-44. [doi: 10.1258/135763307781644960]

## Abbreviations

**CBT:** cognitive behavioral therapy
**DHI:** digital health intervention
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

*Edited by R Kukafka; submitted 07.04.22; peer-reviewed by H Mehdizadeh, R Poss-Doering, L Guo; comments to author 28.06.22; revised version received 12.07.22; accepted 11.08.22; published 26.09.22.*

*Please cite as:*
*Shah A, Hussain-Shamsy N, Strudwick G, Sockalingam S, Nolan RP, Seto E*
*Digital Health Interventions for Depression and Anxiety Among People With Chronic Conditions: Scoping Review*
*J Med Internet Res 2022;24(9):e38030*
*URL: https://www.jmir.org/2022/9/e38030*
*doi:10.2196/38030*
*PMID:36155409*

Review

# Effectiveness of eHealth Self-management Interventions in Patients With Heart Failure: Systematic Review and Meta-analysis

Siru Liu[1*], PhD; Jili Li[2*], BS; Ding-yuan Wan[2], BS; Runyi Li[3], BCS; Zhan Qu[2], BS; Yundi Hu[4], BDS; Jialin Liu[5], MD

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
[2]West China School of Medicine, Sichuan University, Chengdu, China
[3]College of Computer Science, Sichuan University, Chengdu, China
[4]School of Data Science, Fudan University, Shanghai, China
[5]Department of Medical Informatics, West China Hospital, Sichuan University, Chengdu, China
[*]these authors contributed equally

**Corresponding Author:**
Jialin Liu, MD
Department of Medical Informatics
West China Hospital
Sichuan University
No 37 Guoxue Road
Chengdu, 610041
China
Phone: 86 02885422306
Email: dljl8@163.com

## Abstract

**Background:** Heart failure (HF) is a common clinical syndrome associated with substantial morbidity, a heavy economic burden, and high risk of readmission. eHealth self-management interventions may be an effective way to improve HF clinical outcomes.

**Objective:** The aim of this study was to systematically review the evidence for the effectiveness of eHealth self-management in patients with HF.

**Methods:** This study included only randomized controlled trials (RCTs) that compared the effects of eHealth interventions with usual care in adult patients with HF using searches of the EMBASE, PubMed, CENTRAL (Cochrane Central Register of Controlled Trials), and CINAHL databases from January 1, 2011, to July 12, 2022. The Cochrane Risk of Bias tool (RoB 2) was used to assess the risk of bias for each study. The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) criteria were used to rate the certainty of the evidence for each outcome of interest. Meta-analyses were performed using Review Manager (RevMan v.5.4) and R (v.4.1.0 x64) software.

**Results:** In total, 24 RCTs with 9634 participants met the inclusion criteria. Compared with the usual-care group, eHealth self-management interventions could significantly reduce all-cause mortality (odds ratio [OR] 0.83, 95% CI 0.71-0.98, $P$=.03; GRADE: low quality) and cardiovascular mortality (OR 0.74, 95% CI 0.59-0.92, $P$=.008; GRADE: moderate quality), as well as all-cause readmissions (OR 0.82, 95% CI 0.73-0.93, $P$=.002; GRADE: low quality) and HF-related readmissions (OR 0.77, 95% CI 0.66-0.90, $P$<.001; GRADE: moderate quality). The meta-analyses also showed that eHealth interventions could increase patients' knowledge of HF and improve their quality of life, but there were no statistically significant effects. However, eHealth interventions could significantly increase medication adherence (OR 1.82, 95% CI 1.42-2.34, $P$<.001; GRADE: low quality) and improve self-care behaviors (standardized mean difference –1.34, 95% CI –2.46 to –0.22, $P$=.02; GRADE: very low quality). A subgroup analysis of primary outcomes regarding the enrolled population setting found that eHealth interventions were more effective in patients with HF after discharge compared with those in the ambulatory clinic setting.

**Conclusions:** eHealth self-management interventions could benefit the health of patients with HF in various ways. However, the clinical effects of eHealth interventions in patients with HF are affected by multiple aspects, and more high-quality studies are needed to demonstrate effectiveness.

*(J Med Internet Res 2022;24(9):e38697)* doi:10.2196/38697

**KEYWORDS**

heart failure; eHealth; self-management; systematic review; cardiology; cardiovascular; morbidity

## Introduction

Heart failure (HF), a major global public health concern, is a common clinical syndrome caused by cardiac structural or functional impairment [1]. The global prevalence of HF was estimated at 64.34 million cases, and the global economic burden of HF was roughly calculated at US $346.17 billion. Global expenditure related to HF is expected to increase to approximately US $400 billion by 2030 [2]. Since HF cannot be completely cured, it has a major impact on quality of life and requires long-term management [3-5]. The management of patients with HF is a complex issue. Even when the patient is clinically stable, the quality of life is reduced due to dyspnea, depression, fatigue, and cognitive impairment [6]. The symptom burden of patients with HF prevents them from maintaining adequate social life and roles, participating in social activities, and maintaining relationships [7].

Self-management is a dynamic, iterative process in which patients need to employ multidimensional strategies to meet their self-needs for coping with chronic illness in their daily lives [8,9]. Self-management is an effective way to improve the outcome of chronic disease and an important part of the treatment of patients with chronic diseases [10]. Successful self-management requires the active participation of individuals, families, and health care providers [11]. Self-management has been demonstrated to improve clinical outcomes and health-related quality of life, while reducing health care utilization and costs [12-14].

eHealth, as broadly defined, refers to a variety of information and communication technologies used to deliver health care services [15,16]. The use of eHealth interventions to deliver health care services can reduce or eliminate some of the barriers to face-to-face treatment, improve access to treatment, reduce waiting times, and be more cost-effective than face-to-face interventions [17-19]. The World Health Organization (WHO) states that to improve health and reduce health inequities, a rigorous evaluation of eHealth is necessary to generate evidence and facilitate the appropriate integration and use of technology [20]. In addition, the WHO and the International Telecommunication Union launched the National eHealth Strategy Toolkit, which is a comprehensive and practical guide to help all governments, along with their ministries, departments, and agencies, to adapt to suit their own circumstances, goals, and vision in achieving eHealth [21].

There is a large body of research demonstrating the effectiveness of eHealth self-management interventions for chronic diseases [22,23]. eHealth self-management supports patients' empowerment to better manage HF and improve quality of life [24]. Furthermore, eHealth applications and systems could help to reduce the rehospitalization rate of patients with HF and lower the cost of treatment [25]. However, previous research has not specifically focused on the impact of eHealth in self-management. Although eHealth self-management interventions have been used in patients with HF, the results of some systematic reviews have been inconsistent [4,26]. Furthermore, there is a lack of understanding of the effectiveness and influencing factors of eHealth self-management approaches.

To fill these gaps, the aim of this study was to perform a systematic review and meta-analysis of the effectiveness of eHealth self-management interventions in patients with HF. This review will provide a reference for the clinical application and better understanding of the possible benefits of eHealth self-management interventions for patients with HF. These findings will further identify gaps and inform the development of future eHealth interventions.

## Methods

### Search Strategy

This systematic review and meta-analysis was performed following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [27]. We searched the PubMed, EMBASE, CENTRAL (Cochrane Central Register of Controlled Trials), and CINAHL databases from January 1, 2011, to July 12, 2022, to identify randomized controlled trials (RCTs) that provide eHealth self-management interventions for HF patients. We used combined search terms such as (heart failure OR cardiac failure OR heart decompensation) AND (self-management OR self care OR self-administration OR self-medication) AND (Telemedicine OR mobile health OR eHealth OR m-health) AND (randomized controlled trial OR randomization OR randomly). Screening titles/abstracts and full-text evaluation of the articles were managed in a database created with Rayyan software [28]. The full protocol of this systematic review and meta-analysis has been registered in PROSPERO (CRD42021246973).

### Selection Criteria

Inclusion and exclusion criteria strictly followed the Participants-Intervention-Comparison-Outcome (PICO) framework [29]: (1) participants were defined as adults (aged≥18 years) who had been diagnosed with HF (studies in children or adolescents were excluded); (2) interventions consisted of self-management tools including at least one eHealth component (eg, mobile apps), with traditional interventions that did not use any technical support excluded, such as face-to-face meetings; (3) the comparison was HF patients versus usual care; and the (4) outcomes included patient and process outcomes. Patient outcomes are measures directly related to the disease and were considered as the primary outcomes in our study, including all-cause readmission rate, all-cause mortality, HF readmission, and cardiovascular mortality. Process outcomes are measures related to patients' behavior, which were considered as the secondary outcomes in our study, including quality of life, self-care behaviors, HF knowledge, and medication adherence. In addition, only RCTs or cluster RCTs reporting one or more selected outcomes were included. Reviews, editorials, protocols, and non-English papers were excluded.

### Data Extraction

The following characteristics were extracted from studies that met the inclusion criteria: name of the first author, publication year, country, type of eHealth technologies, target of eHealth intervention, duration of intervention, patient demographics, number of participants, recruitment setting, outcomes, descriptions of the control, and the intervention. All outcome data used for the meta-analysis were extracted by the same two independent reviewers (J Li and DW). The synthesis of the results and data charting were performed independently by two researchers (SL and J Li). Disagreements were discussed and consulted with a third reviewer (J Liu) to reach consensus.

### Risk of Bias Assessment

We used the Cochrane Risk of Bias tool (RoB 2) to assess the risk of bias for each study [30]. This tool assesses 5 domains to address different types of bias: randomization process, deviations from the intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Each component includes a low risk of bias, some concerns, or high risk of bias. Risk of bias assessment was performed independently by two reviewers with consultation of the third reviewer when necessary.

### Certainty of Evidence

The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach was used to rate the certainty of evidence for each outcome of interest [31]. This approach rates the risk of bias, inconsistency, indirectness, imprecision, and other considerations (eg, publication bias) as “high,” “moderate,” “low,” or “very low.” The GRADE assessment was completed with the GRADEpro Guideline Development Tool.

### Statistical Analyses

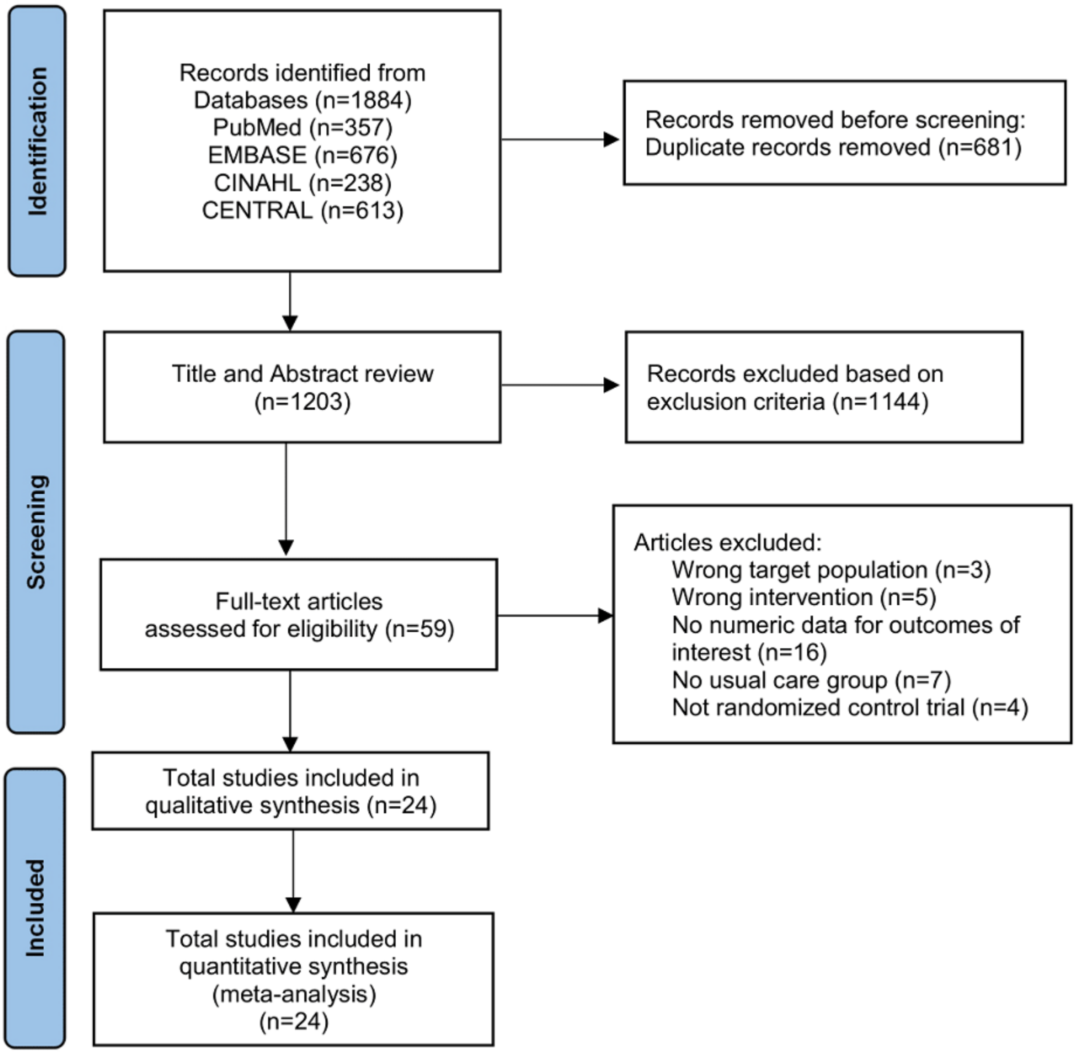
We performed meta-analyses using Review Manager [32] (RevMan v.5.4; Cochrane Training) and R (v.4.1.0 x64). The Mantel-Haenszel odds ratio (OR) was used as the effect size for dichotomous outcomes and the standardized mean difference (SMD) was used to calculate the intervention effect for continuous outcomes. Heterogeneity was then analyzed using the Cochran *Q* test and the $I^2$ statistic. For the *Q* test, $P<.10$ was considered to indicate statistically significant heterogeneity [33]. According to Cochrane Handbook guidelines [34], the interpretation of the $I^2$ statistic was as follows: 0% to 40% may represent not important heterogeneity, 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity, and 75% to 100% may represent considerable heterogeneity. The fixed-effects model was used for outcomes with low heterogeneity and the random-effects model was used when the heterogeneity was significantly high. To measure the publication bias, we constructed funnel plots and performed the Egger test for all primary outcomes to determine the significance of potential asymmetry [35]. The threshold of a significant *P* value was .05. A sensitivity analysis was performed to check the robustness of the pooled results using the leave-one-out approach [36].

## Results

### Search Outcomes

A total of 1884 articles were retrieved in the literature search. After eliminating duplicates, 1203 titles and abstracts were screened in relation to the inclusion/exclusion criteria. Of these, 1144 articles were excluded and a total of 59 articles were subject to full-text review. Finally, 24 articles were included in the systematic review and meta-analysis (Figure 1).

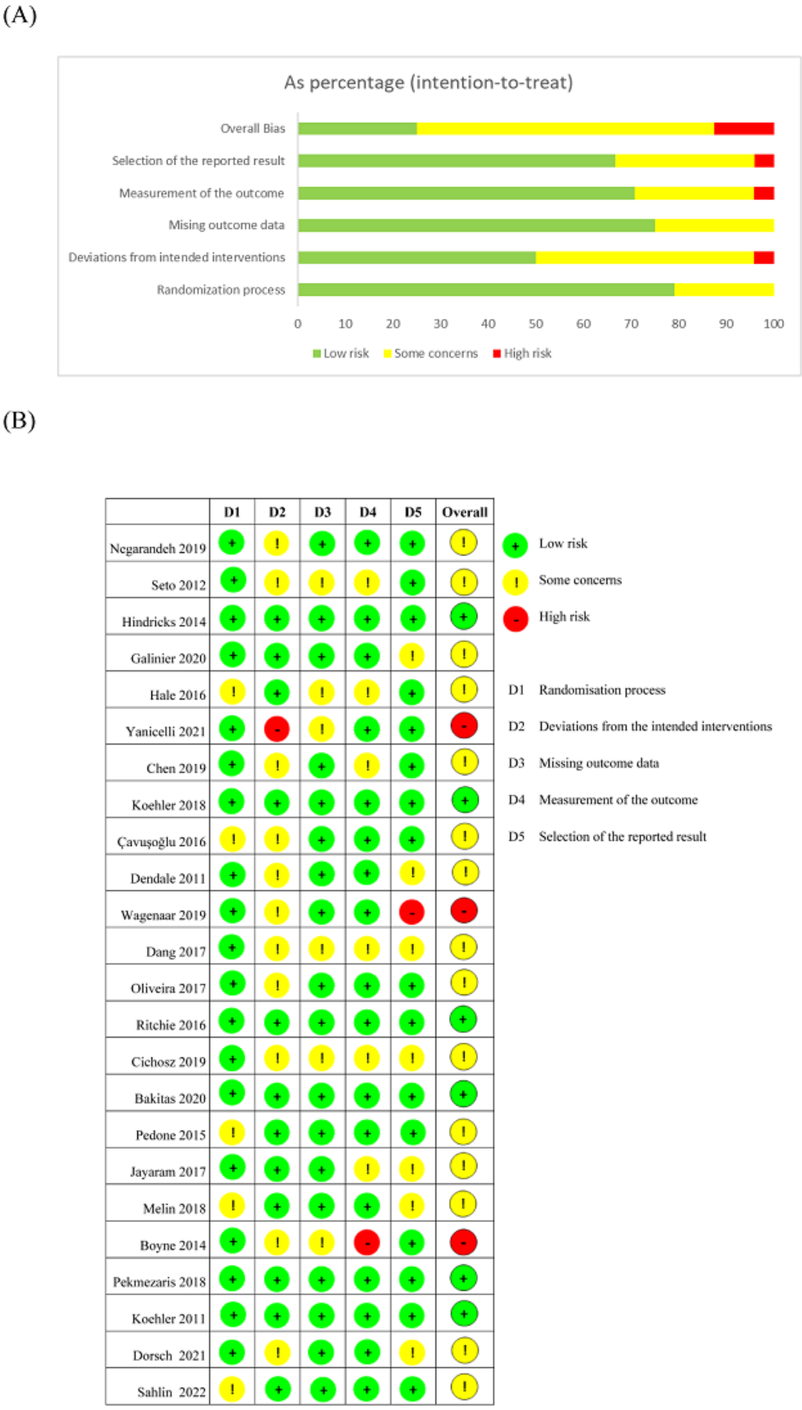
**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for selection and inclusion of the studies via databases. Latest search date: July 12, 2022.



## Risk of Bias and Quality Assessment

The results of risk of bias assessment are summarized in Figure 2. The majority of studies showed a low risk of bias for the randomization process, and only 5 studies were judged to have some concerns due to lack of detailed information about randomization. The studies could not have been blinded to participants considering the nature of the intervention and 11 studies were judged to have some concerns regarding deviations from the intended interventions. One study was determined as "high risk" for the category of measurement of the outcome because all questionnaire scores were patient-reported, which may have possibly affected the outcome [37]. In addition, one study did not report all of the outcomes according to the trial registry record, resulting in a judgment of "high risk" regarding selection of the reported result [38]. For studies included in the meta-analysis, the overall bias was low in 6 studies (25%), with some concerns in 15 studies (63%) and high bias in 3 studies (13%).

**Figure 2.** Quality assessment. (A) Each risk of bias domain presented as a percentage across all included studies. (B) Each risk of bias domain for each included study.



| | D1 | D2 | D3 | D4 | D5 | Overall |
|---|---|---|---|---|---|---|
| Negarandeh 2019 | + | ! | + | + | + | ! |
| Seto 2012 | + | ! | ! | ! | + | ! |
| Hindricks 2014 | + | + | + | + | + | + |
| Galinier 2020 | + | + | + | + | ! | ! |
| Hale 2016 | ! | + | ! | ! | + | ! |
| Yanicelli 2021 | + | - | ! | + | + | - |
| Chen 2019 | + | ! | + | ! | + | ! |
| Koehler 2018 | + | + | + | + | + | + |
| Çavuşoğlu 2016 | ! | ! | + | + | + | ! |
| Dendale 2011 | + | ! | + | + | ! | ! |
| Wagenaar 2019 | + | ! | + | + | - | - |
| Dang 2017 | + | ! | ! | ! | ! | ! |
| Oliveira 2017 | + | ! | + | + | + | ! |
| Ritchie 2016 | + | + | + | + | + | + |
| Cichosz 2019 | + | ! | ! | ! | ! | ! |
| Bakitas 2020 | + | + | + | + | + | + |
| Pedone 2015 | ! | + | + | + | + | ! |
| Jayaram 2017 | + | + | + | ! | ! | ! |
| Melin 2018 | ! | + | + | + | ! | ! |
| Boyne 2014 | + | ! | ! | - | + | - |
| Pekmezaris 2018 | + | + | + | + | + | + |
| Koehler 2011 | + | + | + | + | + | + |
| Dorsch 2021 | + | ! | + | + | ! | ! |
| Sahlin 2022 | ! | + | + | + | + | ! |

+ Low risk
! Some concerns
- High risk

D1 Randomisation process
D2 Deviations from the intended interventions
D3 Missing outcome data
D4 Measurement of the outcome
D5 Selection of the reported result

## Study Characteristics

A total of 9634 participants with HF were included in the 26 (22+2×2) RCTs, with 4820 patients allocated to the eHealth group and 4814 to the usual-care group. Since two studies performed a three-group parallel-randomized trial and reported their results separately [38,39], we divided each of the three-group trials into two RCTs. These studies were performed in 20 different countries, with the largest number of studies performed in the United States (7 studies). Most (15 studies) were multicenter studies, while the other 11 RCTs recruited patients from a single center. The proportion of male HF patients ranged from 46.8% to 83% in the eHealth groups and from 30.2% to 93% in the usual-care groups. The New York Heart Association classification, which classifies patients into one of four categories based on their limitations during physical activity and has been used clinically to determine trial eligibility, was reported in 24 studies. The basic characteristics of the included studies are presented in Table 1, whereas the characteristics of participants and details of interventions are summarized in Multimedia Appendix 1.

**Table 1.** Basic characteristics of studies included in the meta-analysis.

| Reference (year) | Country | Type of eHealth technologies | Target of eHealth intervention | Duration of intervention | Recruitment | Setting |
|---|---|---|---|---|---|---|
| Negarandeh et al [40] (2019) | Iran | Telephone | Education | 2 months | Single center | After discharge |
| Seto et al [41] (2012) | Canada | Designed telemedical system | Monitoring | 6 months | Single center | Ambulatory clinic |
| Hindricks et al [42] (2014) | Australia, Europe[a], and Israel | Designed telemedical system | Monitoring | 12 months | Multicenter (36 sites) | No information |
| Galinier et al [43] (2020) | France | Telephone+designed telemedical system | Monitoring+education | 18 months | Multicenter (38 sites) | After discharge (26.4%),<br>hospitalized (73.6%) |
| Hale et al [44] (2016) | United States | Designed telemedical system | Reminders | 3 months | Multicenter (2 sites) | No information |
| Yanicelli et al [45] (2021) | Argentina | Mobile or tablet app | Monitoring+education | 3 months | Single center | Ambulatory clinic |
| Chen et al [39] (2019)[b] | China | Mobile text message | Education+reminders | 6 months | Single center | After discharge |
| Chen et al [39] (2019)[b] | China | Telephone | Education+reminders | 6 months | Single center | After discharge |
| Koehler et al [46] (2018) | Germany | Designed telemedical system | Monitoring+education | 365-393 days | Multicenter (200 sites) | After discharge |
| Çavuşoğlu et al [47] (2016) | Turkey | Telephone | Education | 6 months | Multicenter (10 sites) | After discharge |
| Dendale et al [48] (2011) | Belgium | Designed telemedical system | Monitoring | 6 months | Multicenter (7 sites) | After discharge |
| Wagenaar et al [38] (2019)[b] | Netherlands | Internet website | Education | 12 months | Multicenter (9 sites) | Ambulatory clinic |
| Wagenaar et al [38] (2019)[b] | Netherlands | Designed telemedical system | Monitoring | 12 months | Multicenter (9 sites) | Ambulatory clinic |
| Dang et al [49] (2017) | United States | Designed telemedical system | Monitoring | 3 months | Single center | Ambulatory clinic |
| Oliveira et al [50] (2017) | Brazil | Telephone | Education | 4 months | Single center | Ambulatory clinic |
| Ritchie et al [51] (2016) | United States | Designed telemedical system | Mixed interventions | 1 months | Single center | After discharge |
| Cichosz et al [52] (2019) | Denmark | Designed telemedical system | Monitoring | 12 months | Multicenter (3 sites) | Ambulatory clinic |
| Bakitas et al [53] (2020) | United States | Telephone | Mixed interventions | 4 months | Multicenter (2 sites) | Ambulatory clinic,<br>hospitalized |
| Pedone et al [54] (2015) | Italy | Telephone+designed telemedical system | Monitoring | 6 months | Single center | Ambulatory clinic<br>after discharge |
| Jayaram et al [55] (2017) | United States | Designed telemedical system | Monitoring | 6 months | Multicenter (33 sites) | After discharge |
| Melin et al [56] (2018) | Sweden | Mobile or tablet app | Monitoring+education | 6 months | Multicenter (3 sites) | After discharge, hospitalized |
| Boyne et al [37] (2014) | Netherlands | Designed telemedical system | Monitoring+education | 12 months | Multicenter (3 sites) | Ambulatory clinic |
| Pekmezaris et al [57] (2018) | United States | Designed telemedical system | Monitoring | 3 months | Single center | After discharge |

| Reference (year) | Country | Type of eHealth technologies | Target of eHealth intervention | Duration of intervention | Recruitment | Setting |
|---|---|---|---|---|---|---|
| Koehler et al [58] (2011) | Germany | Designed telemedical system | Monitoring | Median 26 months (range 12-28 months) | Multicenter (165 sites) | Ambulatory clinic |
| Dorsch et al [59] (2021) | United States | Mobile or tablet app | Mixed interventions | 3 months | Single center | After discharge, hospitalized |
| Sahlin et al [60] (2022) | Sweden | Designed telemedical system | Monitoring+education | 8 months | Multicenter (7 sites) | Ambulatory clinic |

[a] Europe includes only Austria, Czech Republic, Denmark, Germany, and Latvia.

[b] Two types of eHealth interventions employed in a three-group parallel randomized controlled trial design.

### Intervention

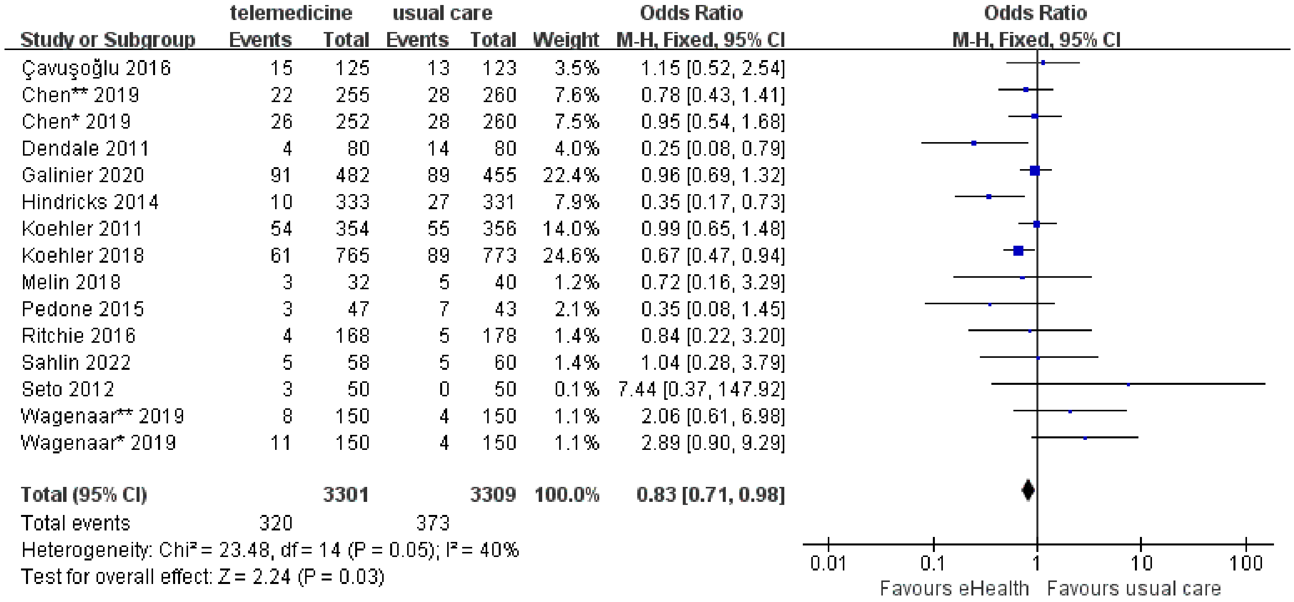
These studies used various types of eHealth technologies for the intervention (Table 1). Of these, 14 studies used a designed telemedical system; 5 studies used the telephone; 3 studies used a mobile phone or tablet app; 2 studies used an internet website and text messages, respectively; and 2 studies used a combination of telephone with a designed telemedical system intervention. The duration of the intervention ranged from 1 month to 28 months (17 studies had interventions lasting≥6 months). The target of the intervention was classified into three types: education, monitoring, and reminders. The mixed targets were defined as the presence of the three types in one study. Educational interventions were defined as interventions that aimed at improving the HF knowledge available through educational programs and instructions. Reminder interventions were defined as interventions that prompted patients to do something (eg, take medication on time) and could be delivered through a designed telemedical system, mobile text messages, or telephone calls. Monitoring interventions were defined as interventions that transmitted vital signs or symptoms (eg, weight, blood pressure, or heart rate) to an external telemedicine center, which may lead to lasting improvements in behaviors. The eHealth intervention varied regarding intensity and the extent of the human component, although most trials provided the eHealth intervention every day (Multimedia Appendix 1).

### Primary Outcomes

#### *Effects of eHealth Interventions on All-Cause Mortality*

A total of 15 studies reported all-cause mortality data. The 15 studies included 6610 participants and showed moderate heterogeneity (*P*=.05; $I^2$=40%). The analysis showed that all-cause mortality was significantly lower in the eHealth intervention group than in the control group (OR 0.83, 95% CI 0.71-0.98, *P*=.03; GRADE: low quality) (Figure 3).

**Figure 3.** Forest plot of the effects of eHealth interventions on all-cause mortality.



| Study or Subgroup | telemedicine Events | telemedicine Total | usual care Events | usual care Total | Weight | Odds Ratio M-H, Fixed, 95% CI |
|---|---|---|---|---|---|---|
| Çavuşoğlu 2016 | 15 | 125 | 13 | 123 | 3.5% | 1.15 [0.52, 2.54] |
| Chen** 2019 | 22 | 255 | 28 | 260 | 7.6% | 0.78 [0.43, 1.41] |
| Chen* 2019 | 26 | 252 | 28 | 260 | 7.5% | 0.95 [0.54, 1.68] |
| Dendale 2011 | 4 | 80 | 14 | 80 | 4.0% | 0.25 [0.08, 0.79] |
| Galinier 2020 | 91 | 482 | 89 | 455 | 22.4% | 0.96 [0.69, 1.32] |
| Hindricks 2014 | 10 | 333 | 27 | 331 | 7.9% | 0.35 [0.17, 0.73] |
| Koehler 2011 | 54 | 354 | 55 | 356 | 14.0% | 0.99 [0.65, 1.48] |
| Koehler 2018 | 61 | 765 | 89 | 773 | 24.6% | 0.67 [0.47, 0.94] |
| Melin 2018 | 3 | 32 | 5 | 40 | 1.2% | 0.72 [0.16, 3.29] |
| Pedone 2015 | 3 | 47 | 7 | 43 | 2.1% | 0.35 [0.08, 1.45] |
| Ritchie 2016 | 4 | 168 | 5 | 178 | 1.4% | 0.84 [0.22, 3.20] |
| Sahlin 2022 | 5 | 58 | 5 | 60 | 1.4% | 1.04 [0.28, 3.79] |
| Seto 2012 | 3 | 50 | 0 | 50 | 0.1% | 7.44 [0.37, 147.92] |
| Wagenaar** 2019 | 8 | 150 | 4 | 150 | 1.1% | 2.06 [0.61, 6.98] |
| Wagenaar* 2019 | 11 | 150 | 4 | 150 | 1.1% | 2.89 [0.90, 9.29] |
| **Total (95% CI)** | | **3301** | | **3309** | **100.0%** | **0.83 [0.71, 0.98]** |
| Total events | 320 | | 373 | | | |

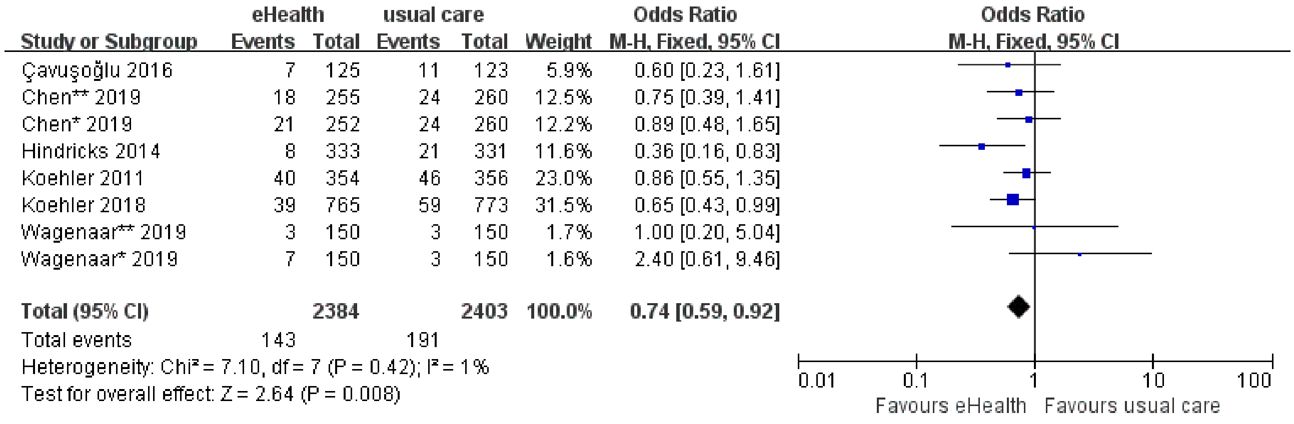
Heterogeneity: Chi² = 23.48, df = 14 (P = 0.05); I² = 40%

Test for overall effect: Z = 2.24 (P = 0.03)

#### *Effects of eHealth Interventions on Cardiovascular Mortality*

A total of 8 studies reported cardiovascular mortality data. The 8 studies included 4787 participants and had low heterogeneity (*P*=.42; $I^2$=1%). The analysis showed that cardiovascular mortality was significantly lower in the eHealth intervention group than in the control group (OR 0.74, 95% CI 0.59-0.92, *P*=.008; GRADE: moderate quality) (Figure 4).

**Figure 4.** Forest plot of the effects of eHealth interventions on cardiovascular mortality.



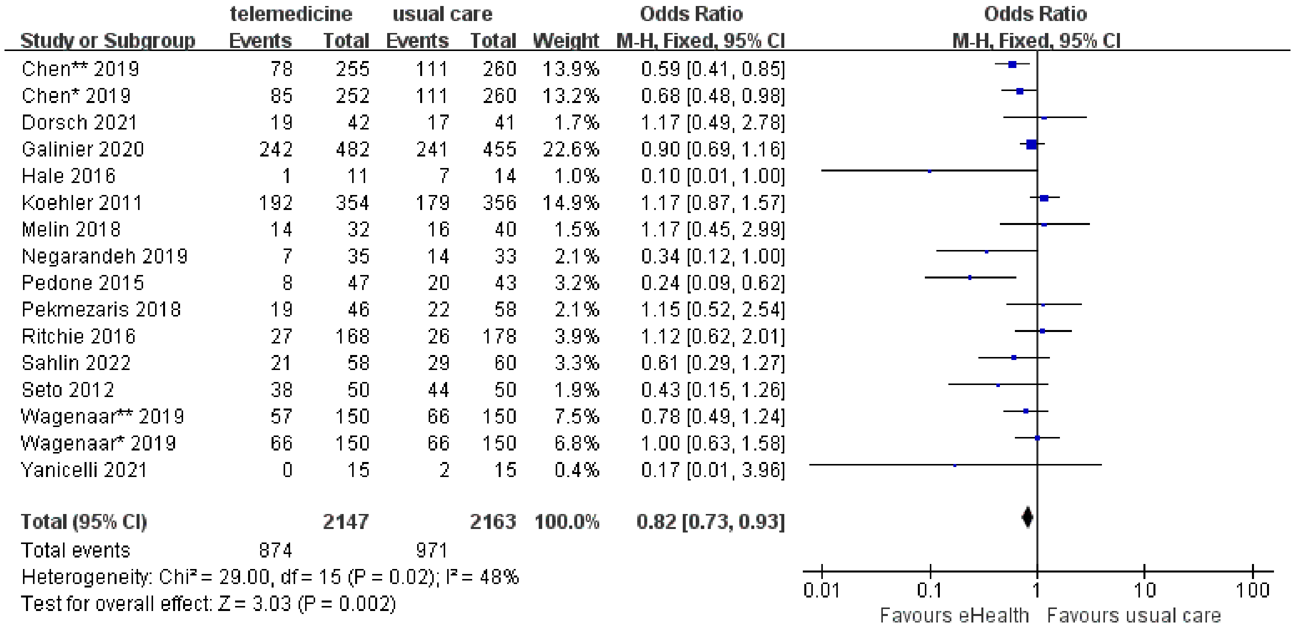
| Study or Subgroup | eHealth Events | eHealth Total | usual care Events | usual care Total | Weight | Odds Ratio M-H, Fixed, 95% CI |
|---|---|---|---|---|---|---|
| Çavuşoğlu 2016 | 7 | 125 | 11 | 123 | 5.9% | 0.60 [0.23, 1.61] |
| Chen** 2019 | 18 | 255 | 24 | 260 | 12.5% | 0.75 [0.39, 1.41] |
| Chen* 2019 | 21 | 252 | 24 | 260 | 12.2% | 0.89 [0.48, 1.65] |
| Hindricks 2014 | 8 | 333 | 21 | 331 | 11.6% | 0.36 [0.16, 0.83] |
| Koehler 2011 | 40 | 354 | 46 | 356 | 23.0% | 0.86 [0.55, 1.35] |
| Koehler 2018 | 39 | 765 | 59 | 773 | 31.5% | 0.65 [0.43, 0.99] |
| Wagenaar** 2019 | 3 | 150 | 3 | 150 | 1.7% | 1.00 [0.20, 5.04] |
| Wagenaar* 2019 | 7 | 150 | 3 | 150 | 1.6% | 2.40 [0.61, 9.46] |
| **Total (95% CI)** | | **2384** | | **2403** | **100.0%** | **0.74 [0.59, 0.92]** |
| Total events | 143 | | 191 | | | |

Heterogeneity: Chi² = 7.10, df = 7 (P = 0.42); I² = 1%
Test for overall effect: Z = 2.64 (P = 0.008)

### *Effects of eHealth Interventions on All-Cause Readmission*

A total of 16 studies reported all-cause readmission data. The 16 studies included 4310 participants and had moderate heterogeneity (*P*=.02; $I^2$=48%). The analysis showed that all-cause readmission was significantly lower in the eHealth intervention group than in the control group (OR 0.82, 95% CI 0.73-0.93, *P*=.002; GRADE: low quality) (Figure 5).

**Figure 5.** Forest plot of the effects of eHealth interventions on all-cause readmission.



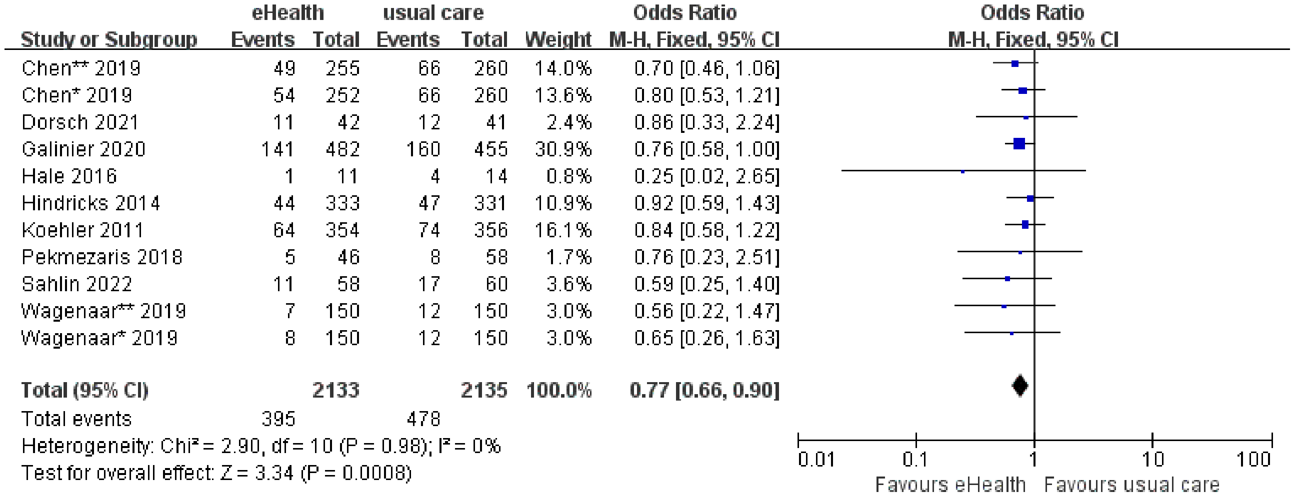
| Study or Subgroup | telemedicine Events | telemedicine Total | usual care Events | usual care Total | Weight | Odds Ratio M-H, Fixed, 95% CI |
|---|---|---|---|---|---|---|
| Chen** 2019 | 78 | 255 | 111 | 260 | 13.9% | 0.59 [0.41, 0.85] |
| Chen* 2019 | 85 | 252 | 111 | 260 | 13.2% | 0.68 [0.48, 0.98] |
| Dorsch 2021 | 19 | 42 | 17 | 41 | 1.7% | 1.17 [0.49, 2.78] |
| Galinier 2020 | 242 | 482 | 241 | 455 | 22.6% | 0.90 [0.69, 1.16] |
| Hale 2016 | 1 | 11 | 7 | 14 | 1.0% | 0.10 [0.01, 1.00] |
| Koehler 2011 | 192 | 354 | 179 | 356 | 14.9% | 1.17 [0.87, 1.57] |
| Melin 2018 | 14 | 32 | 16 | 40 | 1.5% | 1.17 [0.45, 2.99] |
| Negarandeh 2019 | 7 | 35 | 14 | 33 | 2.1% | 0.34 [0.12, 1.00] |
| Pedone 2015 | 8 | 47 | 20 | 43 | 3.2% | 0.24 [0.09, 0.62] |
| Pekmezaris 2018 | 19 | 46 | 22 | 58 | 2.1% | 1.15 [0.52, 2.54] |
| Ritchie 2016 | 27 | 168 | 26 | 178 | 3.9% | 1.12 [0.62, 2.01] |
| Sahlin 2022 | 21 | 58 | 29 | 60 | 3.3% | 0.61 [0.29, 1.27] |
| Seto 2012 | 38 | 50 | 44 | 50 | 1.9% | 0.43 [0.15, 1.26] |
| Wagenaar** 2019 | 57 | 150 | 66 | 150 | 7.5% | 0.78 [0.49, 1.24] |
| Wagenaar* 2019 | 66 | 150 | 66 | 150 | 6.8% | 1.00 [0.63, 1.58] |
| Yanicelli 2021 | 0 | 15 | 2 | 15 | 0.4% | 0.17 [0.01, 3.96] |
| **Total (95% CI)** | | **2147** | | **2163** | **100.0%** | **0.82 [0.73, 0.93]** |
| Total events | 874 | | 971 | | | |

Heterogeneity: Chi² = 29.00, df = 15 (P = 0.02); I² = 48%
Test for overall effect: Z = 3.03 (P = 0.002)

### *Effects of eHealth Interventions on HF-Related Readmission*

A total of 11 studies reported HF-related readmission data. The 11 studies included 4268 participants and had no heterogeneity (*P*=.98; $I^2$=0%). The analysis showed that HF-related readmission was significantly lower in the eHealth intervention group than in the control group (OR 0.77, 95% CI 0.66-0.90, *P*<.001; GRADE: moderate quality) (Figure 6).

**Figure 6.** Forest plot of the effects of eHealth interventions on heart failure–related readmission.

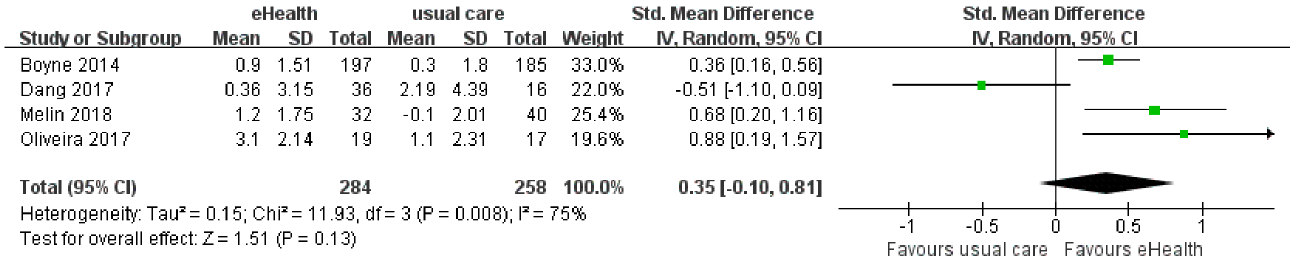


## Secondary Outcomes

### *Effects of eHealth Interventions on HF Knowledge*

A total of 4 studies reported HF knowledge data, 3 of which used the 15-item Dutch HF Knowledge Scale (DHFKS) [61]. One study used a modified questionnaire containing 14 questions, which has proven to be an adequate tool to evaluate the knowledge of Brazilian HF patients [62]. Higher scores across both scales indicate that the patients have more knowledge of HF. The analysis demonstrated an improvement in HF knowledge among patients in the eHealth intervention group compared with that in the control group. However, there was no significant difference between the eHealth and usual-care groups (SMD 0.35, 95% CI –0.10 to 0.81, *P*=.13; GRADE: very low quality), with significantly high heterogeneity between studies (*P*=.008; $I^2$=75%) (Figure 7).

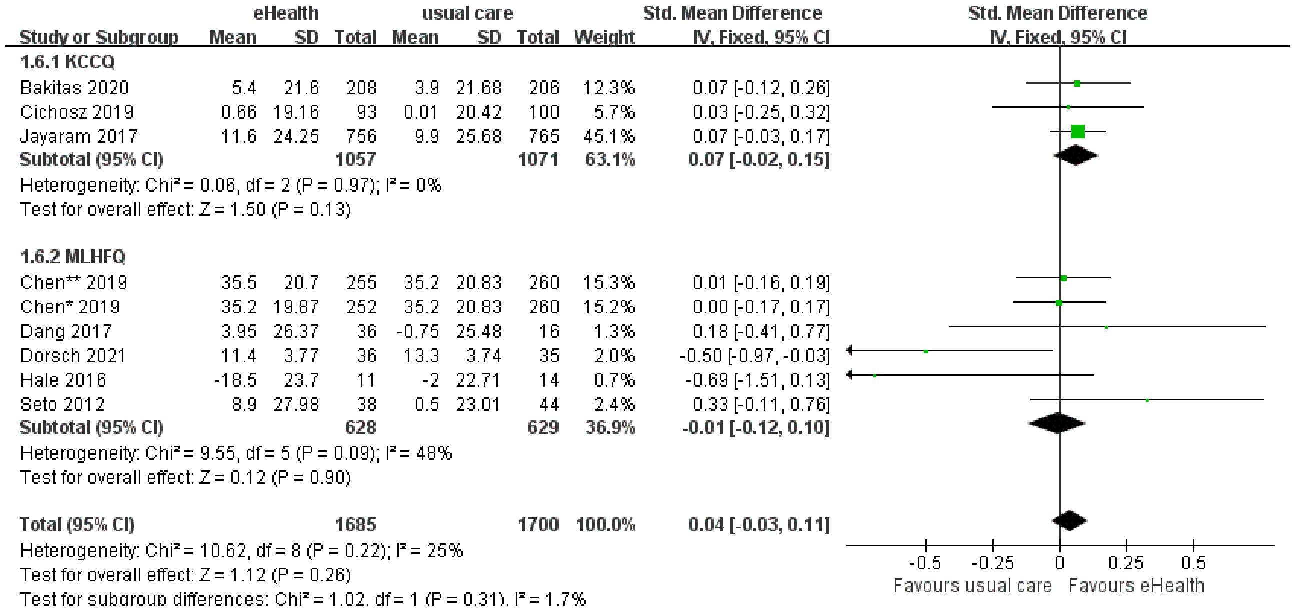
**Figure 7.** Forest plot of the effects of eHealth interventions on heart failure knowledge.



### *Effects of eHealth Interventions on Quality of Life*

A total of 9 studies reported general quality of life data. Three of these studies used the HF-specific 23-item Kansas City Cardiomyopathy Questionnaire (KCCQ), in which higher scores indicate better quality of life [63]. Five of these studies used the Minnesota Living with Heart Failure Questionnaire (MLHFQ), which consists of 21 questions using a 6-point Likert scale [64]. Since lower scores of the MLHFQ indicate a higher quality of life, we calculated the change using the baseline score minus the final score. The analysis demonstrated an improvement in quality of life among patients in the eHealth intervention group compared with that of the control group. However, there was no significant difference between the eHealth and usual-care groups (SMD 0.04, 95% CI –0.03 to 0.11, *P*=.26; GRADE: moderate quality), with low heterogeneity among the 9 studies (*P*=.22; $I^2$=25%) (Figure 8).

**Figure 8.** Forest plot of the effects of eHealth interventions on quality of life. KCCQ: Kansas City Cardiomyopathy Questionnaire; MLHFQ: Minnesota Living with Heart Failure Questionnaire.

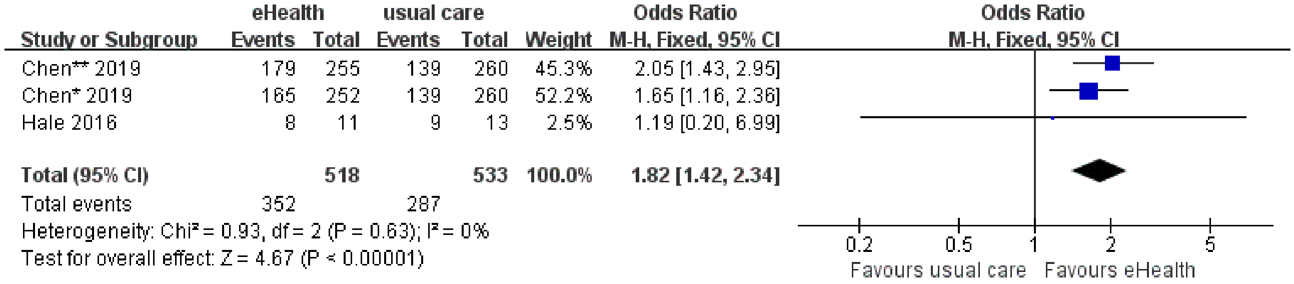


| Study or Subgroup | eHealth Mean | eHealth SD | eHealth Total | usual care Mean | usual care SD | usual care Total | Weight | Std. Mean Difference IV, Fixed, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **1.6.1 KCCQ** | | | | | | | | |
| Bakitas 2020 | 5.4 | 21.6 | 208 | 3.9 | 21.68 | 206 | 12.3% | 0.07 [-0.12, 0.26] |
| Cichosz 2019 | 0.66 | 19.16 | 93 | 0.01 | 20.42 | 100 | 5.7% | 0.03 [-0.25, 0.32] |
| Jayaram 2017 | 11.6 | 24.25 | 756 | 9.9 | 25.68 | 765 | 45.1% | 0.07 [-0.03, 0.17] |
| **Subtotal (95% CI)** | | | **1057** | | | **1071** | **63.1%** | **0.07 [-0.02, 0.15]** |
| Heterogeneity: Chi$^2$ = 0.06, df = 2 (P = 0.97); $I^2$ = 0% | | | | | | | | |
| Test for overall effect: Z = 1.50 (P = 0.13) | | | | | | | | |
| **1.6.2 MLHFQ** | | | | | | | | |
| Chen** 2019 | 35.5 | 20.7 | 255 | 35.2 | 20.83 | 260 | 15.3% | 0.01 [-0.16, 0.19] |
| Chen* 2019 | 35.2 | 19.87 | 252 | 35.2 | 20.83 | 260 | 15.2% | 0.00 [-0.17, 0.17] |
| Dang 2017 | 3.95 | 26.37 | 36 | -0.75 | 25.48 | 16 | 1.3% | 0.18 [-0.41, 0.77] |
| Dorsch 2021 | 11.4 | 3.77 | 36 | 13.3 | 3.74 | 35 | 2.0% | -0.50 [-0.97, -0.03] |
| Hale 2016 | -18.5 | 23.7 | 11 | -2 | 22.71 | 14 | 0.7% | -0.69 [-1.51, 0.13] |
| Seto 2012 | 8.9 | 27.98 | 38 | 0.5 | 23.01 | 44 | 2.4% | 0.33 [-0.11, 0.76] |
| **Subtotal (95% CI)** | | | **628** | | | **629** | **36.9%** | **-0.01 [-0.12, 0.10]** |
| Heterogeneity: Chi$^2$ = 9.55, df = 5 (P = 0.09); $I^2$ = 48% | | | | | | | | |
| Test for overall effect: Z = 0.12 (P = 0.90) | | | | | | | | |
| **Total (95% CI)** | | | **1685** | | | **1700** | **100.0%** | **0.04 [-0.03, 0.11]** |
| Heterogeneity: Chi$^2$ = 10.62, df = 8 (P = 0.22); $I^2$ = 25% | | | | | | | | |
| Test for overall effect: Z = 1.12 (P = 0.26) | | | | | | | | |
| Test for subgroup differences: Chi$^2$ = 1.02, df = 1 (P = 0.31), $I^2$ = 1.7% | | | | | | | | |

## Effects of eHealth Interventions on Medication Adherence

A total of 3 studies reported medication adherence data. The analysis demonstrated a significant improvement in medication adherence among patients in the eHealth intervention group compared with the control group (OR 1.82, 95% CI 1.42-2.34, $P<.001$; GRADE: low quality), with no heterogeneity among 3 studies ($P=.63$; $I^2=0\%$) (Figure 9).

**Figure 9.** Forest plot of the effects of eHealth interventions on medication adherence.

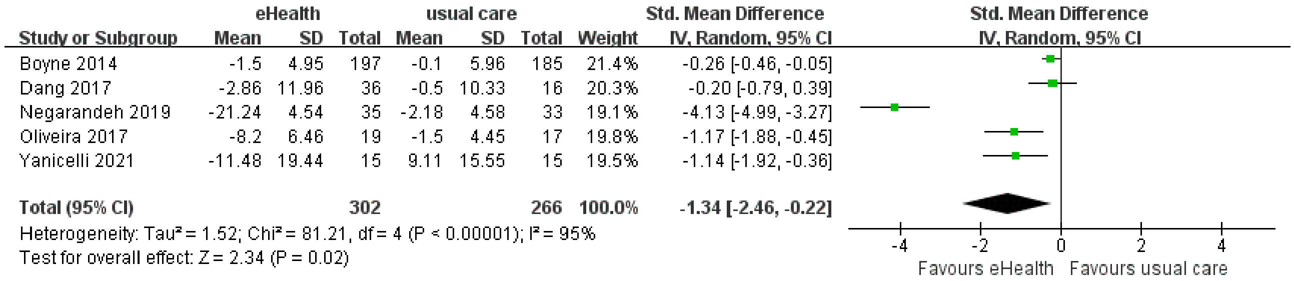


| Study or Subgroup | eHealth Events | eHealth Total | usual care Events | usual care Total | Weight | Odds Ratio M-H, Fixed, 95% CI |
|---|---|---|---|---|---|---|
| Chen** 2019 | 179 | 255 | 139 | 260 | 45.3% | 2.05 [1.43, 2.95] |
| Chen* 2019 | 165 | 252 | 139 | 260 | 52.2% | 1.65 [1.16, 2.36] |
| Hale 2016 | 8 | 11 | 9 | 13 | 2.5% | 1.19 [0.20, 6.99] |
| **Total (95% CI)** | | **518** | | **533** | **100.0%** | **1.82 [1.42, 2.34]** |
| Total events | 352 | | 287 | | | |
| Heterogeneity: Chi$^2$ = 0.93, df = 2 (P = 0.63); $I^2$ = 0% | | | | | | |
| Test for overall effect: Z = 4.67 (P < 0.00001) | | | | | | |

## Effects of eHealth Interventions on Self-care Behaviors

A total of 3 studies reported self-care behavior data. Four of these studies used the European Heart Failure Self-Care Behavior Scale (EHFSC) [65]. Two of these studies used the EHFSC revised to a 9-item scale (EHFScB-9) [66], and 2 used the self-care of heart failure index (SCHFI), which comprises three subscales [67]. Three of the 8 studies had imperfect data and therefore were not included in this meta-analysis. The analysis demonstrated a significant improvement in self-care behaviors among patients in the eHealth intervention group compared with the control group (SMD –1.34, 95% CI –2.46 to –0.22, $P=.02$; GRADE: very low quality), with significantly high heterogeneity among 5 studies ($P<.001$; $I^2=95\%$) (Figure 10).

**Figure 10.** Forest plot of the effects of eHealth interventions on self-care behaviors.



| Study or Subgroup | eHealth Mean | eHealth SD | eHealth Total | usual care Mean | usual care SD | usual care Total | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| Boyne 2014 | -1.5 | 4.95 | 197 | -0.1 | 5.96 | 185 | 21.4% | -0.26 [-0.46, -0.05] |
| Dang 2017 | -2.86 | 11.96 | 36 | -0.5 | 10.33 | 16 | 20.3% | -0.20 [-0.79, 0.39] |
| Negarandeh 2019 | -21.24 | 4.54 | 35 | -2.18 | 4.58 | 33 | 19.1% | -4.13 [-4.99, -3.27] |
| Oliveira 2017 | -8.2 | 6.46 | 19 | -1.5 | 4.45 | 17 | 19.8% | -1.17 [-1.88, -0.45] |
| Yanicelli 2021 | -11.48 | 19.44 | 15 | 9.11 | 15.55 | 15 | 19.5% | -1.14 [-1.92, -0.36] |
| **Total (95% CI)** | | | **302** | | | **266** | **100.0%** | **-1.34 [-2.46, -0.22]** |
| Heterogeneity: Tau$^2$ = 1.52; Chi$^2$ = 81.21, df = 4 (P < 0.00001); $I^2$ = 95% | | | | | | | | |
| Test for overall effect: Z = 2.34 (P = 0.02) | | | | | | | | |

### Subgroup Analyses

Since the inclusion criteria of the trials were different, which contributed to the heterogeneity of studies, subgroup analyses of primary outcomes regarding the enrolled population setting (after discharge or ambulatory clinic) were performed and the results are summarized in Multimedia Appendix 2. Compared with patients in an ambulatory clinic setting, the eHealth intervention showed a larger effect in patients with HF after discharge on the reduction of all-cause mortality (OR 0.73, 95% CI 0.58-0.93, *P*=.01), cardiovascular mortality (OR 0.71, 95% CI 0.53-0.95, *P*=.02), all-cause readmission (OR 0.70, 95% CI 0.56-0.87, *P*=.001), and HF-related readmission (OR 0.75, 95% CI 0.56-1.00, *P*=.05) than the usual-care group. The results showed that the effects of an eHealth intervention may vary in certain population settings.

### Sensitivity Analysis and Publication Bias

Sensitivity analysis was performed for the primary outcomes of meta-analyses, including all-cause mortality, cardiovascular mortality, all-cause readmission rate, and HF-related hospitalizations, using the leave-one-out approach. The direction and magnitude of the combined estimates did not vary markedly with the exclusion of individual studies, indicating the reliability of the findings (Multimedia Appendix 3). The results showed no publication bias in the primary outcomes. No publication bias was detected for any of the primary outcomes. The funnel plots and linear regression test results of funnel plot asymmetry are shown in Multimedia Appendix 4.

## Discussion

### General Findings

There is a reasonable amount of original research exploring the effects of eHealth interventions in HF. However, among the 1884 articles we identified, only 24 met the inclusion criteria for this systematic review. The pooled results suggest that eHealth self-management interventions can improve primary and secondary outcomes in patients with HF.

### Implementation Challenges

#### *Overview*

Although eHealth self-management interventions have the potential to improve chronic disease management, successful implementation in routine clinical practice is rare [68,69]. The main challenges in implementing and utilizing eHealth include user acceptance, standards and interoperability, regulations, ensuring cost-effectiveness and sustainability, and the organization and implementation environment, as discussed in turn below.

#### *eHealth User Acceptance*

End-user acceptance of eHealth products is the key to a successful eHealth program, which is influenced by many factors, although eHealth usability and user training seem to be more important. eHealth products need to improve the understanding of the design and development process from the user's perspective, and focus on developing detailed representations of the user's needs and wants [70-72]. Furthermore, the process of eHealth interventions is complex, which requires active participation, cooperation, and familiarity with the associated processes by all involved [11]. Through eHealth self-management, health care providers can connect more closely with patients, identify problems earlier, provide guidance, and improve patient compliance.

#### *Standards and Interoperability*

A successful eHealth self-management implementation requires unified integration with electronic health record (EHR) and clinical workflows to enable secure and fast access to patient data and information at various locations [73]. This requires interoperability within EHRs, as well as interoperability between EHR and eHealth self-management systems. Ultimately, intelligent eHealth self-management could reduce phone calls and paperwork by automating data exchange. However, eHealth self-management and EHR systems often lack interoperability and standards for clinical data exchange [74,75].

#### *eHealth Regulations*

There is also a lack of eHealth regulations, which is a major barrier to the development of eHealth [76]. The most common regulations used in eHealth implementation are national data and privacy protection laws and regulations, as well as national EHR and health financing legislation [77]. Health financing legislation is critical to ensure the continued sustainability and support of eHealth. National health service laws also provide guidelines for the use and implementation of eHealth [78].

#### *Cost-Effectiveness and Sustainability*

The cost-effectiveness and sustainability of eHealth are key factors that directly affect the successful implementation of eHealth in practice [70]. Many of the costs relate to the development of the eHealth product, deployment, training, product iterations, services, or ongoing maintenance costs of the technology. eHealth self-management programs should ensure cost-effectiveness and sustainability of the program.

#### *Organization and Implementation Environment*

The hospital's management capacity, human resources, and implementation environment are key obstacles to the successful implementation of eHealth [79]. In addition, eHealth technologies also need to fit the organization. The information literacy of patients and easy access to high-speed internet are conducive to the implementation of eHealth [80,81]. eHealth applications should be implemented on top of existing organization-centric and process-controlled systems [82]. This will benefit the efficiency and effectiveness of eHealth solutions and health care services, as well as support the development process and change in management.

### eHealth Self-management in HF

eHealth self-management is a rapidly growing area of HF management, and considerable research is still needed to promote the widespread use of eHealth self-management and enhance its clinical outcomes. Future research should investigate how to expand the content of eHealth self-management tools while adapting to different HF patients. There is a need for a cost-benefit analysis of eHealth self-management, and how to most effectively integrate eHealth self-management into workflows and deploy it in different settings. Further work is needed to understand how eHealth self-management can help with the transformation of care delivery models and how to combine eHealth self-management with a workflow-oriented quality improvement program.

## Limitations

Although this systematic review and associated meta-analyses showed that eHealth self-management interventions benefit patients with HF, this study has several limitations. First, the included studies varied in participant demographics, types of HF, sample size, assessment tools, period of interventions, and type of eHealth interventions, which can lead to heterogeneity and bias. Furthermore, reliance on self-reports and the use of questionnaires to score outcomes may lead to bias in self-reported data. Therefore, the results of this systematic review should be interpreted with caution. Second, participants were recruited at a single site or concentrated in a single population in some cases; thus, the resulting sample was not representative of the entire HF patient population, which would limit the generalizability of the findings. Finally, we only searched for literature in English, excluding other languages. This may limit the retrieval of non-English but related papers.

## Conclusion

In this systematic review, we performed a literature search and provided a comprehensive overview of eHealth self-management in patients with HF. We observed that eHealth self-management can support the health of HF patients in many ways. However, the effectiveness of eHealth in the self-management of patients with HF is affected by multiple aspects and more clinical studies are needed to prove its effectiveness.

## Authors' Contributions

J Liu and SL conceived the study. J Li, SL, J Liu, DW, RL, ZQ, and YH performed the analysis, interpreted the results, and drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Basic characteristics of the studies.
[XLSX File (Microsoft Excel File), 17 KB - jmir_v24i9e38697_app1.xlsx ]

Multimedia Appendix 2
Subgroup analyses for primary outcomes.
[DOCX File , 88 KB - jmir_v24i9e38697_app2.docx ]

Multimedia Appendix 3
Sensitivity analysis of the influence of each study on the pooled estimates.
[DOCX File , 1310 KB - jmir_v24i9e38697_app3.docx ]

Multimedia Appendix 4
Funnel plots of all-cause mortality.
[DOCX File , 48 KB - jmir_v24i9e38697_app4.docx ]

## References

1. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, ESC Scientific Document Group. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure of the European Society of Cardiology (ESC) developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016 Jul 14;37(27):2129-2200. [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]
2. Lippi G, Sanchis-Gomar F. Global epidemiology and future trends of heart failure. AME Med J 2020 Jun 25;5(15):1-6 [FREE Full text]
3. Habal MV, Garan AR. Long-term management of end-stage heart failure. Best Pract Res Clin Anaesthesiol 2017 Jun;31(2):153-166 [FREE Full text] [doi: 10.1016/j.bpa.2017.07.003] [Medline: 29110789]
4. Morken IM, Storm M, Søreide JA, Urstad KH, Karlsen B, Husebø AML. Posthospitalization follow-up of patients with heart failure using eHealth solutions: restricted systematic review. J Med Internet Res 2022 Feb 15;24(2):e32946 [FREE Full text] [doi: 10.2196/32946] [Medline: 35166680]
5. Nordfonn OK, Morken IM, Bru LE, Larsen AI, Husebø AML. Burden of treatment in patients with chronic heart failure - a cross-sectional study. Heart Lung 2021;50(3):369-374 [FREE Full text] [doi: 10.1016/j.hrtlng.2021.02.003] [Medline: 33618147]

6. Pressler SJ, Gradus-Pizlo I, Chubinski SD, Smith G, Wheeler S, Sloan R, et al. Family caregivers of patients with heart failure: a longitudinal study. J Cardiovasc Nurs 2013;28(5):417-428. [doi: 10.1097/JCN.0b013e3182563877] [Medline: 22760173]
7. Olano-Lizarraga M, Wallström S, Martín-Martín J, Wolf A. Interventions on the social dimension of people with chronic heart failure: a systematic review of randomized controlled trials. Eur J Cardiovasc Nurs 2022 Jun 23:online ahead of print. [doi: 10.1093/eurjcn/zvac051] [Medline: 35737922]
8. Miller WR, Lasiter S, Bartlett Ellis R, Buelow JM. Chronic disease self-management: a hybrid concept analysis. Nurs Outlook 2015 Mar;63(2):154-161 [FREE Full text] [doi: 10.1016/j.outlook.2014.07.005] [Medline: 25241136]
9. Ryan P, Sawin KJ. The individual and family self-management theory: background and perspectives on context, process, and outcomes. Nurs Outlook 2009;57(4):217-225 [FREE Full text] [doi: 10.1016/j.outlook.2008.10.004] [Medline: 19631064]
10. Lin C, Hwang S. Patient-centered self-management in patients with chronic kidney disease: challenges and implications. Int J Environ Res Public Health 2020 Dec 16;17(24):9443 [FREE Full text] [doi: 10.3390/ijerph17249443] [Medline: 33339300]
11. Durosier Mertilus DS, Lengacher CA, Rodriguez CS. A review and conceptual analysis of cancer pain self-management. Pain Manag Nurs 2022 Apr;23(2):168-173. [doi: 10.1016/j.pmn.2021.04.005] [Medline: 34006496]
12. Ko D, Bratzke LC, Roberts T. Self-management assessment in multiple chronic conditions: a narrative review of literature. Int J Nurs Stud 2018 Jul;83:83-90. [doi: 10.1016/j.ijnurstu.2018.04.009] [Medline: 29709734]
13. van Beest W, Boon W, Andriessen D, Moors E, van der Veen G, Pol H. Successful implementation of self-management health innovations. J Public Health 2020 Jul 01;30(3):721-735. [doi: 10.1007/s10389-020-01330-y]
14. van Riel PL, Zuidema RM, Vogel C, Rongen-van Dartel SA. Patient self-management and tracking: a European experience. Rheum Dis Clin North Am 2019 May;45(2):187-195 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.008] [Medline: 30952392]
15. Borrelli B, Ritterband LM. Special issue on eHealth and mHealth: challenges and future directions for assessment, treatment, and dissemination. Health Psychol 2015 Dec;34S:1205-1208. [doi: 10.1037/hea0000323] [Medline: 26651461]
16. Kelly M, Fullen B, Martin D, McMahon S, McVeigh JG. eHealth interventions to support self-management in people with musculoskeletal disorders: a scoping review protocol. JBI Evid Synth 2021 Mar;19(3):709-720. [doi: 10.11124/JBIES-20-00147] [Medline: 33725715]
17. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344. [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
18. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in electronic health record patient portal use in nephrology clinics. Clin J Am Soc Nephrol 2015 Nov 06;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
19. Mauco KL, Scott RE, Mars M. Validation of an e-health readiness assessment framework for developing countries. BMC Health Serv Res 2020 Jun 23;20(1):575 [FREE Full text] [doi: 10.1186/s12913-020-05448-3] [Medline: 32576174]
20. World Health Organization. WHO guideline: recommendations on digital interventions for health system strengthening. Geneva, Switzerland: World Health Organization; 2019.
21. World Health Organization, International Telecommunication Union. National eHealth strategy toolkit. World Health Organization. 2012. URL: http://handle.itu.int/11.1002/pub/8069793a-en [accessed 2022-01-12]
22. Butler S, Sculley D, Santos D, Fellas A, Gironès X, Singh-Grewal D, et al. Effectiveness of eHealth and mHealth interventions supporting children and young people living with juvenile idiopathic arthritis: systematic review and meta-analysis. J Med Internet Res 2022 Feb 02;24(2):e30457 [FREE Full text] [doi: 10.2196/30457] [Medline: 35107431]
23. Villarreal V, Berbey-Alvarez A. Evaluation of mHealth applications related to cardiovascular diseases: a systematic review. Acta Inform Med 2020 Jun;28(2):130-137 [FREE Full text] [doi: 10.5455/aim.2020.28.130-137] [Medline: 32742066]
24. Schmaderer M, Miller JN, Mollard E. Experiences of using a self-management mobile app among individuals with heart failure: qualitative study. JMIR Nurs 2021 Aug 09;4(3):e28139 [FREE Full text] [doi: 10.2196/28139] [Medline: 34406966]
25. de la Torre Díez I, Garcia-Zapirain B, Méndez-Zorrilla A, López-Coronado M. Monitoring and follow-up of chronic heart failure: a literature review of eHealth applications and systems. J Med Syst 2016 Jul;40(7):179. [doi: 10.1007/s10916-016-0537-y] [Medline: 27286984]
26. Akinosun AS, Polson R, Diaz-Skeete Y, De Kock JH, Carragher L, Leslie S, et al. Digital technology interventions for risk factor modification in patients with cardiovascular disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Mar 03;9(3):e21061 [FREE Full text] [doi: 10.2196/21061] [Medline: 33656444]
27. 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]
28. 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]
29. Kloda LA, Boruff JT, Cavalcante AS. A comparison of patient, intervention, comparison, outcome (PICO) to a new, alternative clinical question framework for search skills, search results, and self-efficacy: a randomized controlled trial. J Med Libr Assoc 2020 Apr;108(2):185-194 [FREE Full text] [doi: 10.5195/jmla.2020.739] [Medline: 32256230]

30. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
31. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol 2011 Apr;64(4):380-382. [doi: 10.1016/j.jclinepi.2010.09.011] [Medline: 21185693]
32. Review Manager (RevMan). Cochrane Training.: Nordic Cochrane Centre TTC; 2014. URL: https://training.cochrane.org/online-learning/core-software/revman [accessed 2022-01-12]
33. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
34. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M. Cochrane handbook for systematic reviews of interventions. Hoboken, NJ: John Wiley & Sons; 2019.
35. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
36. Liu M, Luong L, Lachaud J, Edalati H, Reeves A, Hwang SW. Adverse childhood experiences and related outcomes among adults experiencing homelessness: a systematic review and meta-analysis. Lancet Public Health 2021 Nov;6(11):e836-e847 [FREE Full text] [doi: 10.1016/S2468-2667(21)00189-4] [Medline: 34599894]
37. Boyne JJJ, Vrijhoef HJM, Spreeuwenberg M, De Weerd G, Kragten J, Gorgels APM, TEHAF investigators. Effects of tailored telemonitoring on heart failure patients' knowledge, self-care, self-efficacy and adherence: a randomized controlled trial. Eur J Cardiovasc Nurs 2014 Jun;13(3):243-252. [doi: 10.1177/1474515113487464] [Medline: 23630403]
38. Wagenaar KP, Broekhuizen BDL, Jaarsma T, Kok I, Mosterd A, Willems FF, et al. Effectiveness of the European Society of Cardiology/Heart Failure Association website 'heartfailurematters.org' and an e-health adjusted care pathway in patients with stable heart failure: results of the 'e-Vita HF' randomized controlled trial. Eur J Heart Fail 2019 Feb;21(2):238-246. [doi: 10.1002/ejhf.1354] [Medline: 30485612]
39. Chen C, Li X, Sun L, Cao S, Kang Y, Hong L, et al. Post-discharge short message service improves short-term clinical outcome and self-care behaviour in chronic heart failure. ESC Heart Fail 2019 Feb;6(1):164-173. [doi: 10.1002/ehf2.12380] [Medline: 30478888]
40. Negarandeh R, Zolfaghari M, Bashi N, Kiarsi M. Evaluating the effect of monitoring through telephone (tele-monitoring) on self-care behaviors and readmission of patients with heart failure after discharge. Appl Clin Inform 2019 Mar;10(2):261-268 [FREE Full text] [doi: 10.1055/s-0039-1685167] [Medline: 30995686]
41. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
42. Hindricks G, Taborsky M, Glikson M, Heinrich U, Schumacher B, Katz A, IN-TIME study group. Implant-based multiparameter telemonitoring of patients with heart failure (IN-TIME): a randomised controlled trial. Lancet 2014 Aug 16;384(9943):583-590. [doi: 10.1016/S0140-6736(14)61176-4] [Medline: 25131977]
43. Galinier M, Roubille F, Berdague P, Brierre G, Cantie P, Dary P, OSICAT Investigators. Telemonitoring versus standard care in heart failure: a randomised multicentre trial. Eur J Heart Fail 2020 Jun;22(6):985-994. [doi: 10.1002/ejhf.1906] [Medline: 32438483]
44. Hale TM, Jethwani K, Kandola MS, Saldana F, Kvedar JC. A remote medication monitoring system for chronic heart failure patients to reduce readmissions: a two-arm randomized pilot study. J Med Internet Res 2016 Apr 17;18(5):e91 [FREE Full text] [doi: 10.2196/jmir.5256] [Medline: 27154462]
45. Yanicelli LM, Goy CB, González VDC, Palacios GN, Martínez EC, Herrera MC. Non-invasive home telemonitoring system for heart failure patients: a randomized clinical trial. J Telemed Telecare 2020 Jan 23;27(9):553-561. [doi: 10.1177/1357633x19899261]
46. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
47. Çavuşoğlu Y, Zoghi M, Eren M, Bozçalı E, Kozdağ G, Şentürk T, Hit-PoinT Investigators. Post-discharge heart failure monitoring program in Turkey: Hit-PoinT. Anatol J Cardiol 2017 Feb;17(2):107-112 [FREE Full text] [doi: 10.14744/AnatolJCardiol.2016.6812] [Medline: 27488754]
48. Dendale P, De Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar;14(3):333-340. [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]
49. Dang S, Karanam C, Gómez-Marín O. Outcomes of a mobile phone intervention for heart failure in a minority county hospital population. Telemed J E Health 2017 Jun;23(6):473-484. [doi: 10.1089/tmj.2016.0211] [Medline: 28051357]
50. Oliveira JAD, Cordeiro RG, Rocha RG, Guimarães TCF, Albuquerque DCD. Impacto do monitoramento telefônico em pacientes com insuficiência cardíaca: ensaio clínico randomizado. Acta Paul Enferm 2017 Aug;30(4):333-342. [doi: 10.1590/1982-0194201700050]

51. Ritchie CS, Houston TK, Richman JS, Sobko HJ, Berner ES, Taylor BB, et al. The E-Coach technology-assisted care transition system: a pragmatic randomized trial. Transl Behav Med 2016 Sep;6(3):428-437 [FREE Full text] [doi: 10.1007/s13142-016-0422-8] [Medline: 27339715]
52. Cichosz SL, Udsen FW, Hejlesen O. The impact of telehealth care on health-related quality of life of patients with heart failure: Results from the Danish TeleCare North heart failure trial. J Telemed Telecare 2020;26(7-8):452-461. [doi: 10.1177/1357633X19832713] [Medline: 30975047]
53. Bakitas MA, Dionne-Odom JN, Ejem DB, Wells R, Azuero A, Stockdill ML, et al. Effect of an early palliative care telehealth intervention vs usual care on patients with heart failure: the ENABLE CHF-PC randomized clinical trial. JAMA Intern Med 2020 Sep 01;180(9):1203-1213 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2861] [Medline: 32730613]
54. Pedone C, Rossi FF, Cecere A, Costanzo L, Antonelli Incalzi R. Efficacy of a physician-led multiparametric telemonitoring system in very old adults with heart failure. J Am Geriatr Soc 2015 Jun;63(6):1175-1180. [doi: 10.1111/jgs.13432] [Medline: 26031737]
55. Jayaram NM, Khariton Y, Krumholz HM, Chaudhry SI, Mattera J, Tang F, et al. Impact of telemonitoring on health status. Circ Cardiovasc Qual Outcomes 2017 Dec;10(12):e004148 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.117.004148] [Medline: 29237746]
56. Melin M, Hägglund E, Ullman B, Persson H, Hagerman I. Effects of a tablet computer on self-care, quality of life, and knowledge: a randomized clinical trial. J Cardiovasc Nurs 2018;33(4):336-343. [doi: 10.1097/JCN.0000000000000462] [Medline: 29369123]
57. Pekmezaris R, Nouryan CN, Schwartz R, Castillo S, Makaryus AN, Ahern D, et al. A randomized controlled trial comparing telehealth self-management to standard outpatient management in underserved Black and Hispanic patients living with heart failure. Telemed J E Health 2019 Oct 01;25(10):917-925 [FREE Full text] [doi: 10.1089/tmj.2018.0219] [Medline: 30418101]
58. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation 2011 May 03;123(17):1873-1880. [doi: 10.1161/CIRCULATIONAHA.111.018473] [Medline: 21444883]
59. Dorsch MP, Farris KB, Rowell BE, Hummel SL, Koelling TM. The effects of the ManageHF4Life mobile app on patients with chronic heart failure: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 07;9(12):e26185 [FREE Full text] [doi: 10.2196/26185] [Medline: 34878990]
60. Sahlin D, Rezanezad B, Edvinsson M, Bachus E, Melander O, Gerward S. Self-care Management Intervention in Heart Failure (SMART-HF): a multicenter randomized controlled trial. J Card Fail 2022 Jan;28(1):3-12 [FREE Full text] [doi: 10.1016/j.cardfail.2021.06.009] [Medline: 34161807]
61. van der Wal MHL, Jaarsma T, Moser DK, van Veldhuisen DJ. Development and testing of the Dutch Heart Failure Knowledge Scale. Eur J Cardiovasc Nurs 2005 Dec;4(4):273-277. [doi: 10.1016/j.ejcnurse.2005.07.003] [Medline: 16126459]
62. Rabelo ER, Mantovani VM, Aliti GB, Domingues FB. Cross-cultural adaptation and validation of a disease knowledge and self-care questionnaire for a Brazilian sample of heart failure patients. Rev Lat Am Enfermagem 2011;19(2):277-284 [FREE Full text] [doi: 10.1590/s0104-11692011000200008] [Medline: 21584373]
63. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol 2000 Apr;35(5):1245-1255 [FREE Full text] [doi: 10.1016/s0735-1097(00)00531-3] [Medline: 10758967]
64. Riegel B, Moser DK, Glaser D, Carlson B, Deaton C, Armola R, et al. The Minnesota Living With Heart Failure Questionnaire: sensitivity to differences and responsiveness to intervention intensity in a clinical population. Nurs Res 2002;51(4):209-218. [doi: 10.1097/00006199-200207000-00001] [Medline: 12131233]
65. Jaarsma T, Strömberg A, Mårtensson J, Dracup K. Development and testing of the European Heart Failure Self-Care Behaviour Scale. European Journal of Heart Failure 2003 Jun 09;5(3):363-370. [doi: 10.1016/s1388-9842(02)00253-2]
66. Jaarsma T, Arestedt KF, Mårtensson J, Dracup K, Strömberg A. The European Heart Failure Self-care Behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument. Eur J Heart Fail 2009 Jan;11(1):99-105. [doi: 10.1093/eurjhf/hfn007] [Medline: 19147463]
67. Riegel B, Carlson B, Moser DK, Sebern M, Hicks FD, Roland V. Psychometric testing of the self-care of heart failure index. J Card Fail 2004 Aug;10(4):350-360. [doi: 10.1016/j.cardfail.2003.12.001] [Medline: 15309704]
68. Lobban F, Appelbe D, Appleton V, Aref-Adib G, Barraclough J, Billsborough J. Health ServicesDelivery Research. An online supported self-management toolkit for relatives of people with psychosis or bipolar experiences: the IMPART multiple case study. Health Serv Deliv Res 2020 Sep;8(37):1-248. [doi: 10.3310/hsdr08370] [Medline: 32986342]
69. Versluis A, van Luenen S, Meijer E, Honkoop PJ, Pinnock H, Mohr DC, et al. SERIES: eHealth in primary care. Part 4: Addressing the challenges of implementation. Eur J Gen Pract 2020 Dec 07;26(1):140-145 [FREE Full text] [doi: 10.1080/13814788.2020.1826431] [Medline: 33025820]
70. Ahmed B, Dannhauser T, Philip N. A systematic review of reviews to identify key research opportunities within the field of eHealth implementation. J Telemed Telecare 2019 Jun;25(5):276-285. [doi: 10.1177/1357633X18768601] [Medline: 29703127]

71. Petersen LS, Bertelsen P, Bjørnes C. Cooperation and communication challenges in small-scale eHealth development projects. Int J Med Inform 2013 Dec;82(12):e375-e385. [doi: 10.1016/j.ijmedinf.2013.03.008] [Medline: 23664669]
72. Polillo A, Gran-Ruaz S, Sylvestre J, Kerman N. The use of eHealth interventions among persons experiencing homelessness: a systematic review. Digit Health 2021 Feb 02;7:2055207620987066 [FREE Full text] [doi: 10.1177/2055207620987066] [Medline: 33598308]
73. Liu J, Liu S, Zheng T, Bi Y. Physicians' perspectives of telemedicine during the COVID-19 pandemic in China: qualitative survey study. JMIR Med Inform 2021 Jun 01;9(6):e26463 [FREE Full text] [doi: 10.2196/26463] [Medline: 33945493]
74. Bernstam EV, Warner JL, Krauss JC, Ambinder E, Rubinstein WS, Komatsoulis G, et al. Quantitating and assessing interoperability between electronic health records. J Am Med Inform Assoc 2022 Apr 13;29(5):753-760. [doi: 10.1093/jamia/ocab289] [Medline: 35015861]
75. Fadrique LX, Rahman D, Vaillancourt H, Boissonneault P, Donovska T, Morita PP. Overview of policies, guidelines, and standards for active assisted living data exchange: thematic analysis. JMIR Mhealth Uhealth 2020 Jun 22;8(6):e15923 [FREE Full text] [doi: 10.2196/15923] [Medline: 32568090]
76. Marija A, Kristina M. Legal aspects of eHealth development in North Macedonia. Bull St. Petersburg Univ 2021;12(3):660-675. [doi: 10.21638/spbu14.2021.311]
77. Scheibner J, Sleigh J, Ienca M, Vayena E. Benefits, challenges, and contributors to success for national eHealth systems implementation: a scoping review. J Am Med Inform Assoc 2021 Aug 13;28(9):2039-2049 [FREE Full text] [doi: 10.1093/jamia/ocab096] [Medline: 34151990]
78. Metzger M, Durand T, Lallich S, Salamon R, Castets P. The use of regional platforms for managing electronic health records for the production of regional public health indicators in France. BMC Med Inform Decis Mak 2012 Apr 03;12:28 [FREE Full text] [doi: 10.1186/1472-6947-12-28] [Medline: 22471902]
79. Kiberu VM, Mars M, Scott RE. Barriers and opportunities to implementation of sustainable e-Health programmes in Uganda: a literature review. Afr J Prim Health Care Fam Med 2017 May 29;9(1):e1-e10 [FREE Full text] [doi: 10.4102/phcfm.v9i1.1277] [Medline: 28582996]
80. Kujala S, Ammenwerth E, Kolanen H, Ervast M. Applying and extending the FITT framework to identify the challenges and opportunities of successful eHealth services for patient self-management: qualitative interview study. J Med Internet Res 2020 Aug 12;22(8):e17696 [FREE Full text] [doi: 10.2196/17696] [Medline: 32784175]
81. Thouvenot VI, Hardiker N. Facilitating eHealth for all through connecting nurses and the Women Observatory for eHealth. Stud Health Technol Inform 2016;225:777-779. [Medline: 27332338]
82. Sharifi M, Ayat M, Jahanbakhsh M, Tavakoli N, Mokhtari H, Wan Ismail WK. E-health implementation challenges in Iranian medical centers: a qualitative study in Iran. Telemed J E Health 2013 Feb;19(2):122-128. [doi: 10.1089/tmj.2012.0071] [Medline: 23374035]

## Abbreviations

**CENTRAL:** Cochrane Central Register of Controlled Trials
**DHFKS:** Dutch Heart Failure Knowledge Scale
**EHFSC:** European Heart Failure Self-Care Behavior Scale
**EHFScB-9:** 9-item European Heart Failure Self-Care Behavior Scale
**EHR:** electronic health record
**GRADE:** Grading of Recommendations, Assessment, Development and Evaluation
**HF:** heart failure
**KCCQ:** Kansas City Cardiomyopathy Questionnaire
**MLHFQ:** Minnesota Living with Heart Failure Questionnaire
**OR:** odds ratio
**PICO:** Participants-Intervention-Comparison-Outcome
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RCT:** randomized controlled trial
**RoB 2:** Cochrane Risk of Bias tool
**SCHFI:** self-care of heart failure index
**SMD:** standardized mean difference
**WHO:** World Health Organization

*Edited by R Kukafka; submitted 12.04.22; peer-reviewed by P Li, M Dorsch; comments to author 09.07.22; revised version received 02.08.22; accepted 06.09.22; published 26.09.22.*

*Please cite as:*
*Liu S, Li J, Wan DY, Li R, Qu Z, Hu Y, Liu J*
*Effectiveness of eHealth Self-management Interventions in Patients With Heart Failure: Systematic Review and Meta-analysis*
*J Med Internet Res 2022;24(9):e38697*
*URL: https://www.jmir.org/2022/9/e38697*
*doi:10.2196/38697*
*PMID:36155484*

Review

# Trends in Effectiveness of Organizational eHealth Interventions in Addressing Employee Mental Health: Systematic Review and Meta-analysis

Elizabeth Stratton[1,2], DClinPsych; Amit Lampit[3], PhD; Isabella Choi[1,2], ClinPsy, PhD; Hanna Malmberg Gavelin[3,4], PhD; Melissa Aji[1], PhD; Jennifer Taylor[1], PhD; Rafael A Calvo[5], PhD; Samuel B Harvey[6,7,8], PhD; Nick Glozier[1,2], PhD

[1]Central Clinical School, Faculty of Health and Medicine, University of Sydney, Camperdown, Australia
[2]ARC Centre of Excellence for Children and Families over the Life Course, Sydney, Australia
[3]Academic Unit for Psychiatry of Old Age, University of Melbourne, Parkville, Australia
[4]Department of Psychology, Umeå University, Umeå, Sweden
[5]Dyson School of Design Engineering, Imperial College London, London, United Kingdom
[6]School of Psychiatry, University of New South Wales, Sydney, Australia
[7]Black Dog Institute, Sydney, Australia
[8]St George Hospital, Sydney, Australia

**Corresponding Author:**
Rafael A Calvo, PhD
Dyson School of Design Engineering
Imperial College London
Imperial College Rd, South Kensington,
London, SW7 9EG
United Kingdom
Phone: 44 20 7594 888
Email: r.calvo@imperial.ac.uk

## Abstract

**Background:** Mental health conditions are considered the leading cause of disability, sickness absence, and long-term work incapacity. eHealth interventions provide employees with access to psychological assistance. There has been widespread implementation and provision of eHealth interventions in the workplace as an inexpensive and anonymous way of addressing common mental disorders.

**Objective:** This updated review aimed to synthesize the literature on the efficacy of eHealth interventions for anxiety, depression, and stress outcomes in employee samples in organizational settings and evaluate whether their effectiveness has improved over time.

**Methods:** Systematic searches of relevant articles published from 2004 to July 2020 of eHealth intervention trials (app- or web-based) focusing on the mental health of employees were conducted. The quality and bias of all studies were assessed. We extracted means and SDs from publications by comparing the differences in effect sizes (Hedge *g*) in standardized mental health outcomes. We meta-analyzed these data using a random-effects model.

**Results:** We identified a tripling of the body of evidence, with 75 trials available for meta-analysis from a combined sample of 14,747 articles. eHealth interventions showed small positive effects for anxiety (Hedges $g$=0.26, 95% CI 0.13-0.39; $P$<.001), depression (Hedges $g$=0.26, 95% CI 0.19-0.34; $P$<.001), and stress (Hedges $g$=0.25, 95% CI 0.17-0.34; $P$<.001) in employees' after intervention, with similar effects seen at the medium-term follow-up. However, there was evidence of no increase in the effectiveness of these interventions over the past decade.

**Conclusions:** This review and meta-analysis confirmed that eHealth interventions have a small positive impact on reducing mental health symptoms in employees. Disappointingly, we found no evidence that, despite the advances in technology and the enormous resources in time, research, and finance devoted to this area for over a decade, better interventions are being produced. Hopefully, these small effect sizes do not represent optimum outcomes in organizational settings.

**Trial Registration:** PROSPERO CRD42020185859; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=185859

***(J Med Internet Res 2022;24(9):e37776)*** doi:10.2196/37776

**KEYWORDS**

eHealth; mental health; employee; systematic review; mobile phone

## Introduction

Mental health conditions are to be considered the leading cause of disability, sick leave, and long-term work incapacity in most developed countries [1]. Furthermore, poor mental health has a substantial impact on employee well-being, productivity, absenteeism, compensation claims, and social welfare systems [2,3]. Evidence supports the increased demand for workplace interventions, highlighting that working conditions and the workplace environment can influence employees' mental health and well-being [4].

The nature of mental health symptoms fluctuates on a continuum between thriving and struggling [5]. Most mental ill health that is seen in the workforce is because of common mental disorders, most notably, depression, anxiety, and stress-related conditions [1,6,7]. Employees who have a mental health condition and become too unwell to continue working rarely move straight from being *healthy* to needing sick leave [8]. There is usually a course that an employee might experience as they develop worsening symptoms [9], and different interventions may be required at different stages of this course.

International approaches suggest multilevel organizational approaches targeting (1) healthy workers via universal prevention interventions; (2) those with subclinical conditions (symptomatic or at-risk workers), such as those experiencing high stress via indicated interventions; and (3) those workers who have disclosed a mental health condition with tertiary interventions (treatments) [8]. Furthermore, the UK Thriving at Work Review into Mental Health at Work recommends the inclusion of the use of therapeutically tailored interventions based on individual-specific needs [5].

The 2 potential rate-limiting steps for organizations in implementing mental health intervention programs are the budget and logistics of delivering universal interventions to all staff at scale and the ability to target indicated and early interventions to those who are at risk or unwell before disclosure. The internet offers a unique opportunity to address these rate-limiting steps by delivering eHealth interventions with components such as cognitive behavioral therapy (CBT) and stress management to a broad audience. eHealth interventions provide employees with access to psychological help when they are not employed in a typical working environment, such as shift workers or those working from home, or when they may be sick-listed from work. As a result, eHealth interventions have been widely implemented and provided. eHealth is an emerging field in public health and business and provides health services and information delivered or enhanced through the internet and related technologies, such as smartphone apps. Recently, eHealth has been seen as a popular approach in organizations as it provides an inexpensive and anonymous way of addressing common mental disorders [10], including apps linked to wearable devices, and guided meditation programs [11].

There is a plethora of evidence for the short- and long-term benefits of eHealth-delivered CBT for treating anxiety and depressive conditions in both the general population and clinical settings [12-15]. Evidence has also emerged for the effectiveness of mindfulness-based eHealth interventions in improving symptoms for both the general population and individuals who are symptomatic [16].

However, we know that employed individuals differ systemically from both general and clinical populations used in most eHealth studies; for instance, employees have much better mental health (fewer symptoms) than general and clinical populations [17], for whom many eHealth interventions and their content have been developed. As such, there are likely to be floor effects and other efficacy modifiers. The delivery in, and by, organizations will be different from just open access to interventions in the general population. This influences uptake and engagement, which are known determinants of digital health efficacy [18].

A previous meta-analysis considering eHealth interventions from 2004 (first identified) to 2017 [19] found randomized controlled trial (RCT) evidence for only 23 eHealth interventions delivered to employees. Overall, there was a small pooled effect of reducing depression, anxiety, and stress symptoms after the intervention (Hedges $g$=0.24), which was sustained at follow-up (Hedges $g$=0.23). Similar effect sizes (ESs) were found in a contemporaneous review [20]. Only 2 years later, a meta-analysis [21] identified 50% (n=34) more studies of such interventions and suggested stronger effects on stress (Hedges $g$=0.54), insomnia (Hedges $g$=0.70), and burnout (Hedges $g$=0.51) but not symptoms of depression or anxiety.

None of the previous reviews have addressed the key issue of efficacy, which is affected by the *digital placebo effect*. Smartphone app users can experience significantly reduced mental health symptoms, even if the app does not contain any direct therapeutic intervention, similar to the placebo effect seen in pharmaceutical trials when using "active" or "attention" controls [22], or may reflect regression to the mean of fluctuating symptoms. Even simply tracking the symptoms of depression in an app can lead to significant reductions in depression [23]. Both processes result in an apparent "intra-group" effect without any therapeutic intervention. In a trial, the type of control used is a primary determinant of the "between-group" ES. In a recent systematic review of smartphone apps for anxiety in clinical samples, the between-group ESs were lower in trials that used active controls than in those that used passive controls [24]. This inflates the apparent efficacy of the interventions evaluated in trials with passive controls, such as the wait-list. It is not known whether and to what extent this is observed in working populations or for more preventive approaches.

Considering the increasing number of interventions being developed, implemented, and potentially evaluated, this review aimed to address the following questions by systematically reviewing the current state of evidence for the efficacy of eHealth interventions in reducing depression, anxiety, and stress in employees (Textbox 1).

**Textbox 1.** Research questions.

**Research questions addressed by this review**

- Is there evidence for the improvement in effectiveness over time (ie, are the interventions getting better)?
- What factors, if any, moderate the efficacy?
    - Intervention approach; indicated, tailored, tertiary or universal
    - Type of control (active vs wait-list)
    - Type of intervention: cognitive behavioral therapy, mindfulness, stress management, and other
    - Presence of in-person support

# Methods

## Overview

This review complies with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [25]. We aimed to identify all published and unpublished, peer-reviewed, controlled clinical trials of eHealth interventions targeted at employees reporting outcomes on a standardized mental health measure of depression, anxiety, and stress. The systematic review protocol was registered on PROSPERO (International Prospective Register of Systematic Reviews; CRD42020185859).

## Search Strategy

A systematic literature search was conducted using MEDLINE, PsycINFO, Cochrane Register of Controlled Trials, and Embase electronic databases for relevant peer-reviewed articles in English of controlled trials and RCTs published from 2004 (when the first eHealth intervention was identified) [19] to July 2020. Keywords used related to “workplace,” “intervention,” “outcome,” and “study design.” An example of this search strategy is presented in Multimedia Appendix 1. The search terms were developed from our previous systematic review [19]. The tables of contents of the *Journal of Medical Internet Research*, *Journal of Internet Interventions*, *Occupational and Environmental Medicine*, and the *Journal of Occupational and Environmental Medicine*, as well as the reference lists of included studies, were manually searched.

## Study Selection Criteria

### *Eligibility Criteria*

The PICOS (Patient/Population, Intervention, Comparison, and Outcomes) framework commonly used to identify components of clinical evidence for systematic reviews in evidence-based medicine, endorsed by the Cochrane Collaboration [26], was used in this review.

#### Participants

Participants had to be in current paid employment and working-age adults aged between 18 and 65 years. Studies were excluded if the sample was defined as volunteer workers, unemployed participants, or general or clinical populations.

#### Interventions

Any eHealth intervention, defined as a therapeutic intervention delivered through a website, smartphone, tablet, or mobile app and designed to improve mental health, was included.

#### Controls

Studies were required to have a control group, defined as either *passive* (care as usual, no contact, or wait-list) or *active* (eg, another eHealth intervention as the comparison group).

#### Outcomes

The study had to report on at least one common mental disorder outcome: depression, anxiety, or stress. All eligible outcomes for each study and domain were included. The outcomes were standardized mean difference (SMD) from preintervention baseline score to immediate postintervention use and the follow-up time point.

The following measurements of common mental disorders were regarded as eligible:

1. Diagnostic interview
2. Self-report diagnosis by a physician, psychologist, or other qualified health professional
3. Self-administered rating scale for mental health (anxiety, depression, or stress)

#### Studies

Studies were excluded from the review if they were not specifically limited to employees and delivered in a workplace setting or used in-person, telephone, and email interventions only.

### *Identification of Studies*

After duplicates were removed, 3 independent authors (ES, IC, and MA) screened all titles and abstracts to identify potentially relevant studies. Abstracts and full-text versions of potentially eligible studies were independently assessed by 2 investigators (ES and IC). Eligible studies with individual citations were scanned to ensure that all relevant studies were identified. Disagreements were adjudicated in conjunction with the senior author (NG).

### Data Extraction and Coding

We extracted the mean and SD of standardized measures of stress or distress, depression, and anxiety and the sample size (n) in each arm (intervention and control) at baseline and at each follow-up time point. When sufficient details were not reported, the authors were contacted. Additional data concerning the type of design, intervention and control details, participant characteristics, study duration, length of follow-up, organizational population, and general outcomes were recorded (see the *Results* section). Studies were also coded by intervention type: *universal* interventions targeting relatively healthy participants; *indicated* interventions targeting subclinical symptoms or syndromes; and *tertiary* treatments for explicit diagnoses.

Studies reporting results of the same intervention and sample in different papers (eg, a postintervention and follow-up study) were treated as 1 study, and we used the first follow-up outcome point in the analysis.

We categorized studies according to their mental health outcome measures: anxiety, depression, or stress. Studies with multiple outcomes appear in >1 mental health outcome result.

### Quality Assessment and Risk of Bias Within Studies

The risk of bias in RCTs was assessed using the revised Cochrane risk of bias tool for RCTs (RoB version 2.0) [27], in which 5 domains were independently evaluated by 2 authors: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Each domain was assessed for the risk of bias. Studies were graded as (1) "low risk of bias" when a low risk of bias was determined for all domains, (2) "some concerns" if at least one domain was assessed as raising some concerns but not to be at high risk of bias for any single domain, or (3) "high risk of bias" when a high risk of bias was determined for at least one domain or there were bias concerns in multiple domains [27]. If a study contained >1 intervention arm, each intervention was assessed as 1 study.

### Statistical Analysis

The summary ES was the SMD (calculated as Hedge *g*, with 95% CI) between the intervention and control groups for each outcome measure (anxiety, depression, and stress) [28]. A positive ES (SMD) indicated that the intervention was more efficacious than the control. The level of significance was set at $P<.05$ and 95% CIs. The magnitude of the effect was categorized as large (SMD>0.8), moderate (SMD 0.5-0.8), small (SMD 0.2 to <0.5), or trivial (SMD<0.2) [29]. Pooling of ESs across studies was performed using a random-effects model in Comprehensive Meta-Analysis (Biostat Inc; version 3). We detected and accommodated outliers in the meta-analysis using the random-effects variance shift model and the likelihood ratio test. As the different types of eHealth interventions and approaches used may result in large heterogeneity [30], between-study heterogeneity was quantified using $\tau^2$ (variance of true effects) and further assessed using the Higgins $I^2$ statistic, which estimates the proportion of observed variance not because of sampling error. An $I^2$ value of 75% was considered large, 50% was considered moderate, and 25% was considered low [31].

We estimated the results separately for each mental health outcome measure (anxiety, depression, and stress) at the postintervention and first follow-up time points. Meta-regressions were performed for each mental health outcome to identify whether the intervention approach (eg, universal), type of intervention (eg, CBT), type of control (eg, wait-list), or intervention delivery method (web-based vs smartphone) moderated the observed effect.

To evaluate differences in effectiveness over time, we used only within-group ESs from the intervention arms for anxiety, depression, or stress outcomes. ESs were calculated by dividing the difference between pre- and posttreatment means by the pooled SD of the difference [32]. ANOVA was used to assess whether there were any differences between the mean ESs for each year. For a test for trends, the within - group ESs with positive signs indicated improvement, and negative signs indicated worsening. This method allowed for the difference in effects over time to be accurately calculated, allowing us to explore whether interventions improved with technology.

To assess small study effects (in cases where at least 10 studies were available for analysis), we used a funnel plot for the overall effects and each subgroup analysis, which compared the outcome effects with their SEs. We used the Egger regression test to examine further asymmetry of the funnel plot [33] with statistical significance based on $P<.10$. In cases where at least 10 studies were available and a small study effect was found, we used a Duval and Tweedie trim and fill analysis to quantify the magnitude of the small study effect [34].

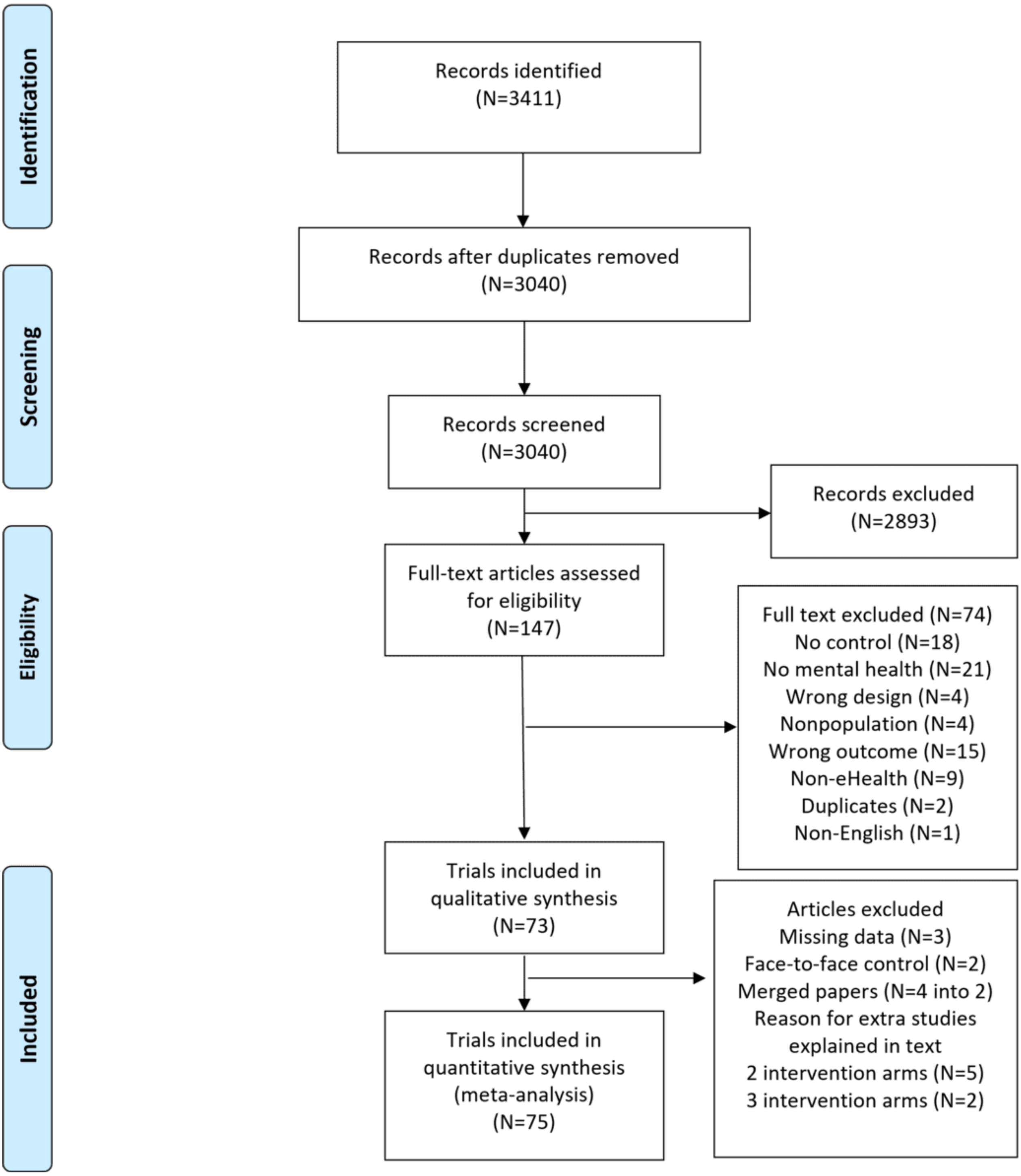
## Results

### Search Results

The search strategy identified 3411 titles (Figure 1). Of the 3411 articles, after the removal of 371 (10.88%) duplicates, 3040 (89.12%) titles and abstracts were reviewed by the authors (ES, IC, and MA). Of these 3040 articles, 2893 (95.16%) articles were excluded based on the eligibility criteria, leaving 147 (4.84%) articles that were potentially relevant to the research question. Their full texts were examined by 2 independent researchers (ES and MA), and discrepancies were decided by the senior author (NG). Of the 147 articles, 74 (50.3%) were excluded for reasons provided in Figure 1. Of the remaining 73 studies, data were missing for 8 (11%) studies. The authors were contacted, and 5 responded with relevant data [35-39]. A total of 3 authors did not respond or were unable to obtain data; thus, these studies were excluded from the meta-analysis [35,40,41]. Of the 73 studies, 4 (5%) articles reported on the same study at different follow-up time points and were merged [42-45], leaving 68 (93%) studies, of which 2 (3%) were excluded as they used face-to-face treatment for control groups [46,47]. The remaining 97% (66/68) of articles were identified as meeting the criteria for quality assessment. Of the 66 studies, 5 (8%) used 2 intervention arms [48-52], and 2 (3%) studies used 3 intervention arms [44,53]. In this case, the intervention arms were treated as individual trials reported as author names;

year of publication; and the letters a, b, or c (eg, Smith [a]). The number of participants in the control group was split evenly as a comparative arm to ensure that participants were not counted twice. This process resulted in 75 trials (Figure 1) for the meta-analysis. Multimedia Appendix 2 [36-39,42-45,49-53] provides the list of references.

**Figure 1.** Flow Diagram of included studies.



## Risk of Bias

The individual studies included in the meta-analysis showed that, overall, 12% (9/75) had a low risk of bias, 3% (2/75) had ≥1 concern, and 85% (64/75) had a high risk of bias. Most of the bias was observed in deviations from the intended interventions and missing outcome data, which is not uncommon and is an area for improvement in the eHealth field. An important aspect common to behavioral interventions is that 55% (41/75) of the studies were rated as having a high risk of bias because of the use of a wait-list control group. As such, it was impossible for users to remain blinded (Multimedia Appendix 3 [36-39,42-45,49-53]).

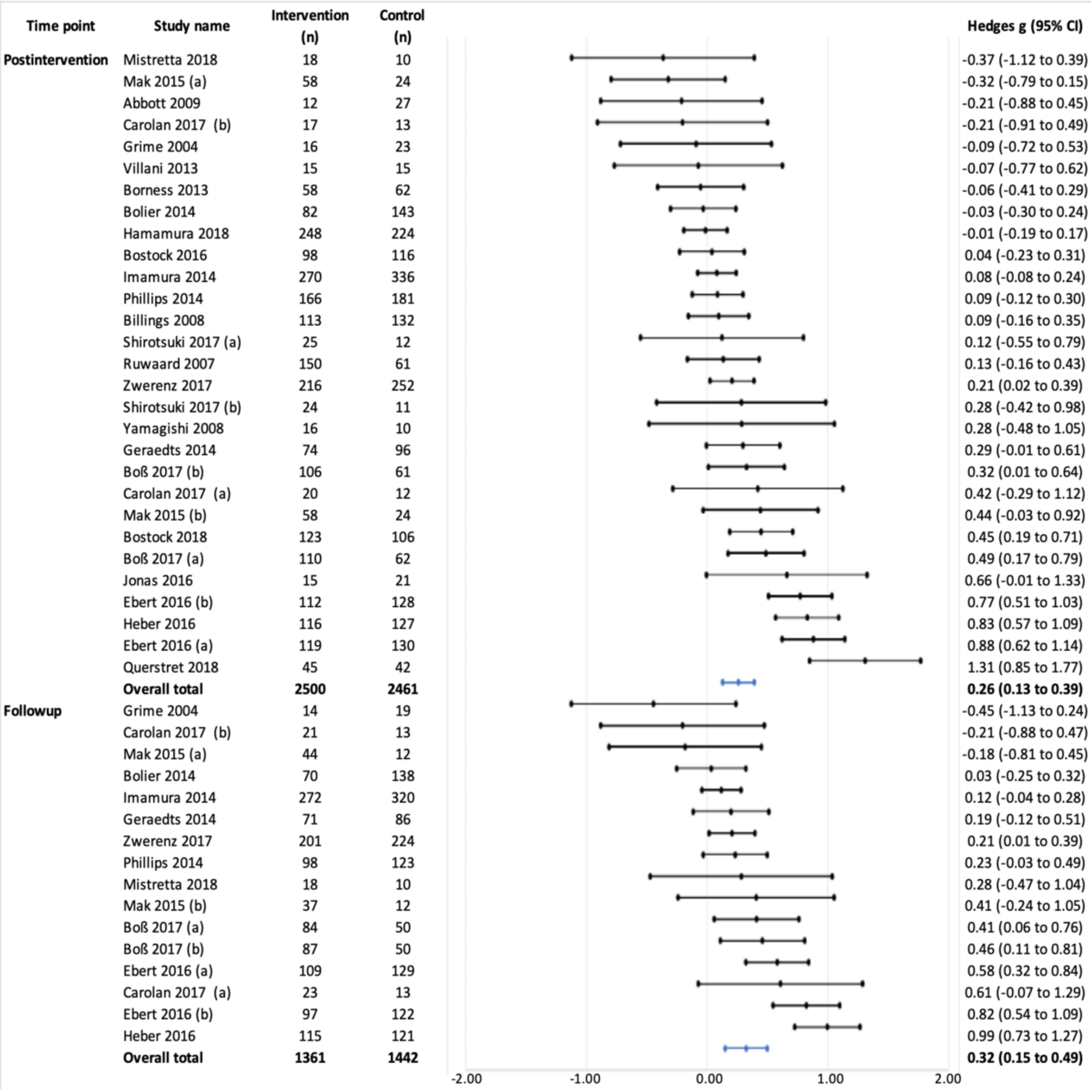
All the included trials used web-based self-report measures to collect the baseline and follow-up data. Multimedia Appendix 4 provides an overview of the scales used and their prevalence.

## Meta-analysis Results

### *Overall Summary of the Identified Interventions*

The 75 trials reported postintervention data for 14,747 participants. Most trials used universal approaches (47/75, 63%), with fewer indicated (10/75, 13%), tertiary (11/75, 15%), or tailored (7/75, 9%) intervention approaches. Most interventions were delivered via web-based platforms as opposed to smartphone apps. Most of the trials used a wait-list control group. The 3 most common types of interventions described by the authors were CBT, mindfulness, and stress management. The studies were conducted in 15 countries: the United States (16/75, 21%), Japan (15/75, 20%), Germany (13/75, 17%), the United Kingdom (7/75, 9%), the Netherlands (5/75, 7%), Australia (4/75, 5%), Sweden (4/75, 5%), China (2/75, 3%), Europe combined (2/75, 3%), Italy (2/75, 3%), Brazil (1/75, 1%), Finland (1/75, 1%), Hong Kong (1/75, 1%), Singapore (1/75, 1%), and the United States and Canada combined (1/75, 1%). The most common types of participants were health care professionals (18/75, 24%). Interventions for the insurance industry (7/75, 9%), managers (6/75, 8%), information technology (6/75, 8%), male-dominated industries (5/75, 7%), telecommunications (5/75, 7%), schools (3/75, 4%), universities (3/75, 4%), marketing and sales (3/75, 4%), banking (1/75, 1%), and human resources (1/75, 1%) were also evaluated (Figure 2). The term *general employee* is used in Multimedia Appendix 2 when the study did not mention the organization type. A total of 15 trials were conducted across several industries, and only 2 were conducted in employees on sick leave. Further summarized and detailed descriptions of the studies are presented in Table 1 and Multimedia Appendix 5 [36-39,42-45,49-53].

**Figure 2.** Effects on anxiety symptoms.



| Time point | Study name | Intervention (n) | Control (n) | Hedges g (95% CI) |
|---|---|---|---|---|
| **Postintervention** | Mistretta 2018 | 18 | 10 | -0.37 (-1.12 to 0.39) |
| | Mak 2015 (a) | 58 | 24 | -0.32 (-0.79 to 0.15) |
| | Abbott 2009 | 12 | 27 | -0.21 (-0.88 to 0.45) |
| | Carolan 2017 (b) | 17 | 13 | -0.21 (-0.91 to 0.49) |
| | Grime 2004 | 16 | 23 | -0.09 (-0.72 to 0.53) |
| | Villani 2013 | 15 | 15 | -0.07 (-0.77 to 0.62) |
| | Borness 2013 | 58 | 62 | -0.06 (-0.41 to 0.29) |
| | Bolier 2014 | 82 | 143 | -0.03 (-0.30 to 0.24) |
| | Hamamura 2018 | 248 | 224 | -0.01 (-0.19 to 0.17) |
| | Bostock 2016 | 98 | 116 | 0.04 (-0.23 to 0.31) |
| | Imamura 2014 | 270 | 336 | 0.08 (-0.08 to 0.24) |
| | Phillips 2014 | 166 | 181 | 0.09 (-0.12 to 0.30) |
| | Billings 2008 | 113 | 132 | 0.09 (-0.16 to 0.35) |
| | Shirotsuki 2017 (a) | 25 | 12 | 0.12 (-0.55 to 0.79) |
| | Ruwaard 2007 | 150 | 61 | 0.13 (-0.16 to 0.43) |
| | Zwerenz 2017 | 216 | 252 | 0.21 (0.02 to 0.39) |
| | Shirotsuki 2017 (b) | 24 | 11 | 0.28 (-0.42 to 0.98) |
| | Yamagishi 2008 | 16 | 10 | 0.28 (-0.48 to 1.05) |
| | Geraedts 2014 | 74 | 96 | 0.29 (-0.01 to 0.61) |
| | Boß 2017 (b) | 106 | 61 | 0.32 (0.01 to 0.64) |
| | Carolan 2017 (a) | 20 | 12 | 0.42 (-0.29 to 1.12) |
| | Mak 2015 (b) | 58 | 24 | 0.44 (-0.03 to 0.92) |
| | Bostock 2018 | 123 | 106 | 0.45 (0.19 to 0.71) |
| | Boß 2017 (a) | 110 | 62 | 0.49 (0.17 to 0.79) |
| | Jonas 2016 | 15 | 21 | 0.66 (-0.01 to 1.33) |
| | Ebert 2016 (b) | 112 | 128 | 0.77 (0.51 to 1.03) |
| | Heber 2016 | 116 | 127 | 0.83 (0.57 to 1.09) |
| | Ebert 2016 (a) | 119 | 130 | 0.88 (0.62 to 1.14) |
| | Querstret 2018 | 45 | 42 | 1.31 (0.85 to 1.77) |
| | **Overall total** | **2500** | **2461** | **0.26 (0.13 to 0.39)** |
| **Followup** | Grime 2004 | 14 | 19 | -0.45 (-1.13 to 0.24) |
| | Carolan 2017 (b) | 21 | 13 | -0.21 (-0.88 to 0.47) |
| | Mak 2015 (a) | 44 | 12 | -0.18 (-0.81 to 0.45) |
| | Bolier 2014 | 70 | 138 | 0.03 (-0.25 to 0.32) |
| | Imamura 2014 | 272 | 320 | 0.12 (-0.04 to 0.28) |
| | Geraedts 2014 | 71 | 86 | 0.19 (-0.12 to 0.51) |
| | Zwerenz 2017 | 201 | 224 | 0.21 (0.01 to 0.39) |
| | Phillips 2014 | 98 | 123 | 0.23 (-0.03 to 0.49) |
| | Mistretta 2018 | 18 | 10 | 0.28 (-0.47 to 1.04) |
| | Mak 2015 (b) | 37 | 12 | 0.41 (-0.24 to 1.05) |
| | Boß 2017 (a) | 84 | 50 | 0.41 (0.06 to 0.76) |
| | Boß 2017 (b) | 87 | 50 | 0.46 (0.11 to 0.81) |
| | Ebert 2016 (a) | 109 | 129 | 0.58 (0.32 to 0.84) |
| | Carolan 2017 (a) | 23 | 13 | 0.61 (-0.07 to 1.29) |
| | Ebert 2016 (b) | 97 | 122 | 0.82 (0.54 to 1.09) |
| | Heber 2016 | 115 | 121 | 0.99 (0.73 to 1.27) |
| | **Overall total** | **1361** | **1442** | **0.32 (0.15 to 0.49)** |

**Table 1.** Summarized description of the selected studies.

| Study | Type | Country and population (sample size) | Intervention and duration (sample size) | Control (sample size) | Type of intervention |
|---|---|---|---|---|---|
| Grime[a] [54] | CBT[b] | United Kingdom; Occupational health department (48) | • “Beating the Blues” web-based program for depression and anxiety<br>• 8 web-based sessions, which last approximately an hour per week (24) | WLC[c] (24) | Tertiary |
| Hasson et al[a,d] [55] | CBT | Sweden; IT and media companies (317) | • 12-month open access<br>• Web-based self-help exercises (129) | WLC+information (174) | Universal |
| Shimazu et al[a] [56] | CBT | Japan; construction machinery company (225) | • 1-month web-based psychoeducation based on social cognitive theory<br>• Self-based program with 3 phases (5 chapters; 112) | WLC (113) | Universal |
| Cook et al[a] [57] | Stress management | United States; human resource employees (419) | • Web-based multimedia health promotion program “Health Connection” (209) | WLC+paper-based information(210) | Universal |
| Ruwaard et al[a,d] [58] | CBT | The Netherlands; general employees (239) | • 7-week web-based program<br>• Supported by trained therapists with 10 personalized feedback sessions<br>• 1 module per week: awareness, relaxation, worrying, positive self-verbalization, positive assertiveness, and time management<br>• In-person and audio training (177) | WLC (62) | Tailored |
| Billings et al[e] [59] | CBT | United States; technology company (309) | • 3-month open-access web-based program (154) | WLC (155) | Tailored |
| Suzuki et al[a] [60] | CBT | Japan; university staff (43) | • A 2-week program with 4 modules<br>• Daily monitoring and feedback and sleep diary<br>• Emailed weekly summary and advice (21) | WLC (22) | Universal |
| Yamagishi et al[a] [61] | Assertion training | Japan; shift working nurses (60) | • 9 weeks of 60-minute web-based training provided weekly (30) | WLC (30) | Universal |
| Abbott et al[a] [62] | CBT | Australia; industrial organization (53) | • Web-based program<br>• 7 core modules (26) | WLC (27) | Universal |
| Bennett et al[a] [63] | Behavior change | United States; general managers (145) | • Web-based open-access ExecuPrev<br>• At least 10 hours over 6 months (72) | WLC (73) | Universal |
| Glück and Maercker[a] [64] | Mindfulness | Austria, Germany, and Switzerland; universities, car dealerships, broadcasting stations, and health care consulting (50) | • Web-based program for 13 days<br>• 2 modules, with each module lasting for 6 days with 20 minutes per day (28) | WLC (21) | Universal |
| Borness et al[a] [65] | Cognitive training | Australia; public sector general employees (135) | • 16 weeks of web-based cognitive training based on memory, attention, language, and executive function<br>• The program was called “Spark!” with three 20-minute sessions per week (67) | Active control program (68); general knowledge information | Universal |
| Feicht et al[a] [66] | Positive psychology | Germany; insurance company (147) | • 7-week web-based happiness training<br>• Weekly modules that took 10 to 15 minutes<br>• Emailed instructions once weekly (85) | WLC (62) | Universal |

| Study | Type | Country and population (sample size) | Intervention and duration (sample size) | Control (sample size) | Type of intervention |
|---|---|---|---|---|---|
| Ketelaar et al [38][a] | Health Surveillance Model | The Netherlands; nurses and allied health professionals (367) | • Tailored, assessed by screening Psyfit: well-being based; healthy participants<br>• Strong at work: stress management; learning skills to cope with work stress<br>• Color your life: depressive symptoms<br>• Don’t Panic Online: panic symptoms for subclinical and mild cases of panic disorder<br>• Drinking less: reducing risky alcohol drinking behavior (178) | WLC (188) | Tailored |
| Lappalainen et al[a,d] [67] | CBT and ACT[f] | Finland; men aged 28 to 58 years with depression (23) | • “P4Well”—a 3-month program, including 3 group meetings, an internet or web portal, mobile phone apps, and personal monitoring devices (11) | WLC (12) | Tertiary |
| Villani et al[a,d] [68] | Stress management | Italy; female oncology nurses with high stress (30) | • 4-week program, with eight 5-minute video clips twice weekly, with a narrative<br>• After work on the study phone (15) | Active control with 8 video clips (15) | Indicated |
| Bolier et al[a,d] [69] | Health Surveillance Model | The Netherlands; nurses, allied health professionals, and general employees (423) | • Tailored, assessed by screening Psyfit: well-being based; healthy participants<br>• Strong at work: stress management; learning skills to cope with work stress<br>• Color your life: depressive symptoms<br>• Don’t Panic Online: panic symptoms for subclinical and mild cases of panic disorder<br>• Drinking less: reducing risky alcohol drinking behavior (212) | WLC (211) | Tailored |
| Deitz et al[a] [70] | Behavior change | United States; hospital employees with cardiovascular risk (210) | • 6-week web-based program with weekly modules<br>• On the basis of increasing knowledge and reducing risk (105) | WLC (105) | Universal |
| Ebert et al[a,d] [10] | Problem-solving training | Germany; teachers (150) | • 6 weeks, 5 lessons<br>• Web-based program with 1 lesson per week and practice between each lesson (75) | WLC (75) | Tertiary |
| Geraedts et al [42,43] (2 papers)[a,d] | Problem-solving training and cognitive therapy | The Netherlands; banking (231) | • Web-based Happy@Work program<br>• 6 weekly sessions, participants had to complete an assessment each week to move on (116) | WLC (115) | Tertiary |
| Imamura et al [44,45] (2 papers)[a,d] | CBT | Japan; IT (762) | • 6-week web-based program<br>• 6 lessons, 1 lesson per week, approximately 30 minutes each<br>• Each lesson had homework (381) | WLC+information (381) | Tertiary |
| Ly et al[a,d] [71] | ACT | Sweden; middle managers in the private sector (73) | • 6-week smartphone app<br>• 6 modules, 1 per week<br>• Audio lecture, text, and exercises supported by student psychologist (36) | WLC (37) | Universal |
| Mori et al[a,d] [72] | CBT | Japan; IT engineers with high computer literacy (168) | • A 4-week program with homework<br>• Web-based<br>• 150-minute group class; web-based entries to log daily stresses (85) | WLC (83) | Universal |
| Phillips et al[a] [73] | CBT | United Kingdom; transport, health, and communication sectors (637) | • 5-week web-based “MoodGYM” program<br>• 5 modules of 1 hour for preventing and coping with depression (318) | WLC+information (319) | Tertiary |
| Umanodan et al[a] [74] | Stress management | Japan; manufacturing company (266) |  | WLC (121) | Universal |

| Study | Type | Country and population (sample size) | Intervention and duration (sample size) | Control (sample size) | Type of intervention |
|---|---|---|---|---|---|
| | | | • 6-week web-based program<br>• 6 lessons (1 per week); self-paced<br>• 2-phased learning process (142) | | |
| Carissoli et al[e] [75] | Mindfulness | Italy; general employees (56) | • 3-week smartphone app<br>• Practice 2 meditations per day, lasting 15 minutes each (20) | Music; 2 songs per day, lasting approximately 15 minutes each, while doing nothing else (18) | Universal |
| Cook et al[a] [76] | Stress management | United States; older IT employees (278) | • “HealthyPast50,” a web-based open-access program<br>• 3 months, 5 modules (138) | WLC (140) | Universal |
| Ebert et al[a] [77] | Behavior change | Germany; teachers (128) | • 6-week internet-based weekly sessions (64) | WLC (64) | Universal |
| Guille et al [36][a] | CBT | United States; medical interns (199) | • MoodGYM program, comprising 4 weekly, web-based sessions lasting approximately 30 minutes each (100) | 4 weekly mental health information emails (99) | Universal |
| Mak et al [51][a] | MBSR[g] | China; university (321) | • 2-arm intervention: 8-week web-based mindfulness training, 1 lesson per week that took 23 to 30 minutes (107)<br>• The second group had the identical training plus health action process approach(107) | WLC (107) | Universal |
| Prasek[a] [78] | Mindfulness | Unite States; university (192) | • 7-week self-guided, web-based mindfulness program called Sherman Project (101) | WLC (91) | Universal |
| Stansfeld et al[a,d] [79] | Stress management | United Kingdom; NHS[h] Mental Health Trust (275) | • Web-based team-based health promotion program<br>• 6 fortnightly modules for 3 months (216) | WLC (59) | Universal |
| Volker et al [39][a] | CBT and PST[i] | The Netherlands; sick-listed (773) | • Return@Work, a web-based program with 5 modules<br>• Up to 17 sessions, ranging from 6 to 17 (131) | WLC (89) | Tertiary |
| Yuan[a] [80] | PST | Hong Kong | • Happy@Work, a web-based program with 4 modules (162) | WLC (159) | Universal |
| Allexandre et al [53][a] | Mindfulness | United States; general employees from a corporate call center (91) | • 8-week web-based program with 1 session per week; audio guided; daily articles available and 2 email reminders sent (54)<br>• Access to the above plus meeting in groups for 1 hour once a week; group session deep breathing exercise for 2 minutes, 10-minute audio recording, 20- to 30-minute guided meditation, and 20 minutes of discussion questions (37)<br>• Same as above plus weeks 3, 6, and 8 were facilitated by a licensed clinical (33) | WLC (37) | Universal |
| Birney et al[a] [81] | CBT | United States (300) | • 6-week CBT mobile phone app “MoodHacker”<br>• Brief daily interactions (150) | WLC (150) | Tertiary |
| Bostock et al[a] [82] | CBT | United States; office based (270) | | WLC (135) | Universal |

| Study | Type | Country and population (sample size) | Intervention and duration (sample size) | Control (sample size) | Type of intervention |
|---|---|---|---|---|---|
| | | | • 8-week web-based Sleepio.com intervention<br>• Animated virtual therapist ("The Prof"), sleep diary data, email or SMS text message prompts, and moderated web-based community (135) | | |
| Dyrbye et al[a] [83] | Positive psychology | United States; practicing physicians (290) | • 6-week web-based program<br>• One 5-minute domain per week (145) | WLC (145) | Universal |
| Ebert et al[a,d] [84] (a) | Stress management | Germany (264) | • 7-week web-based intervention GET.ON Stress<br>• Weekly modules, 45 to 60 minutes each<br>• Daily stress diaries and e-coach (psychologist; 132) | WLC (132) | Indicated |
| Ebert et al[a,d] [85] (b) | Stress management | Germany (264) | • 7-week web-based intervention GET.ON Stress<br>• Weekly modules, 45 to 60 minutes each<br>• Daily stress diaries (132) | WLC (132) | Indicated |
| Heber et al[a,d] [86] | Stress management | Germany (264) | • 7 sessions, with 1 to 2 sessions per week; web-based (132) | WLC (132) | Indicated |
| Hersch et al[a] [87] | Stress management | United States; nurses and nurse managers in public hospitals (104) | • Web-based intervention BREATHE, with open access for 3 months and 7 modules (52) | WLC (52) | Universal |
| Imamura et al[a] [44] | CBT | Japan (1236) | • 4-week internet access to UTSMed, comprising text and illustrations; 90 pages (276)<br>• Same as above but for moderate depression (291)<br>• Same as above but for high depression (51) | WLC no depression (285); same as above but for moderate depression (290); same as above but for high depression (43) | Tailored |
| Jonas et al[a,d] [88] | CBT | Germany; employees with burnout (59) | • A 4-week web-based program "Beratung Hilft" ("counselling helps")<br>• Structured and therapist guided<br>• Daily stress diary (18) | WLC (21) | Indicated |
| Beiwinkel et al[a,d] [89] | CBT | Germany (180) | • "HelpID," a 12-week web-based program with weekly sessions of 30 to 45 minutes<br>• Weekly reminder emails (100) | WLC+information (80) | Tertiary |
| Boß et al [48] (a)[a] and (b)[a,d] | CBT | Germany (434) | • 5-week web-based program with 1 module per week (146)<br>• Identical to the GET.ON intervention plus additional adherence-focused guidance by e-coaches (trained psychologist; 144) | WLC (144) | Universal |
| Carolan et al [49][a,d] | CBT | United Kingdom (84) | • WorkGuru, a web-based program with 7 weekly modules (28)<br>• Same as above plus weekly web-based, guided discussion group (28) | WLC (28) | Indicated |
| Shirotsuki et al [52][a] | CBT | Japan; office workers in hospitality and sales company (87) | • 6-week web-based program where participants watched weekly e-learning movie segments (5-10 minutes long); recorded daily mood and weekly homework (29)<br>• Same as above plus participants consumed 1 bottle of the supplement soft drink (29) | Recorded their mood state every day on a weekly monitoring sheet (homework; 29) | Universal |

| Study | Type | Country and population (sample size) | Intervention and duration (sample size) | Control (sample size) | Type of intervention |
|---|---|---|---|---|---|
| Zhang et al[a] [90] | Awareness training | Singapore; health care workers (80) | • 4-week smartphone-based program<br>• 20-minute briefing session and a daily SMS text messaging service (40) | Information from the Health Promotion Board (40) | Universal |
| Zwerenz et al[a,d] [91] | Psychoeducation | Germany; sick-listed and in inpatient rehabilitation (652) | • 12-week web-based program with weekly reflective 45-minute blogs instructed by a therapist<br>• Reminder emails, individualized feedback, audio-guided stress management, homework, and forum (303) | Active control; regular email reminders to use selected information posted on the web about stress management and coping (329) | Tertiary |
| Bostock et al[a] [92] | Mindfulness | United Kingdom; pharmaceutical and high-tech employees (238) | • 8-week Headspace smartphone app<br>• 45 days of daily 10- to 20-minute mindfulness meditation (128) | NHS web-based advice for work stress (110) | Universal |
| Eriksson et al[a] [93] | Mindful self-compassion program | Sweden; practicing psychologists (101) | • 6 weeks web-based program<br>• Videos with guided instructions (51) | WLC (49) | Universal |
| Gollwitzer et al [50][a] | Mental contrasting | Germany; general nurses (129) | • Web-based 3-week program; daily written reflections on “what if”; identifying obstacles in the way (41)<br>• Same as above plus structured daily time for use (41) | WLC (47) | Universal |
| Hamamura et al[e] [94] | CBT | Japan; marketing company (557) | • A 4-week smartphone app called “jibun kiroku”<br>• Daily activities on an hourly basis<br>• Evaluate the quality of their sleep, mood, and energy level (306) | WLC (251) | Indicated |
| Imamura et al [37][a,d] | CBT | Japan; telecommunications (706) | • 6-week 6-lesson web-based program with 30-minute lessons per week<br>• Voluntary homework and reminder emails (353) | WLC (353) | Universal |
| Lilly et al[a] [95] | MBSR | United States and Canada; emergency telecommunication (323) | • 7-week web-based (Destress 9-1-1) program, with 1 module per week for 30 minutes; introduction video, texts, and a moderated discussion board; outside practice for up to 45 minutes of daily mindfulness homework<br>• Guided audio (163) | WLC (160) | Universal |
| Mistretta et al[a] [96] | MBRT[j] | United States; hospital employees (38) | • 6-week smartphone app (23) | WLC (15) | Indicated |
| Oishi et al[a] [97] | CBT | Japan; teachers (240) | • 12-week web-based “Mind Skill Up Training” program<br>• 7 modules plus 1 group session, including homework, mood tracking, audio, and visual narrator, with 6 reminder emails (120) | WLC (120) | Universal |
| Persson Asplund et al[a,d] [98] | Stress management | Sweden; middle managers in health care, education, IT, or communications sectors (117) | • 8 weekly modules<br>• 2 to 3 hours per week to complete<br>• Personalized written feedback via email from a coach (psychologist; 59) | Active control; weekly mail contact, homework, and access to a moderated discussion forum with other users (58) | Indicated |

| Study | Type | Country and population (sample size) | Intervention and duration (sample size) | Control (sample size) | Type of intervention |
|---|---|---|---|---|---|
| Querstret et al[a] [99] | MBCT[k] | United Kingdom (118) | - 4-week web-based program (60) | WLC (58) | Universal |
| Song et al[a] [100] | Health Surveillance Model | Japan (1526) | - 16-week smartphone app “Karada-no-ki-mochi”<br>- Records weekly and daily moods (612) | WLC (914) | Universal |
| Coelhoso et al[a,d] [101] | Mindfulness meditation | Brazil; female private hospital employees (490) | - 8-week smartphone app 4 with classes per week<br>- Each class contained a brief theoretical portion and a 15-minute guided practice<br>- Participants wrote reflections in a gratitude journal for 20 minutes per week<br>- Pop-ups, notifications, and feedback scores (250) | Active control; similar program (240) | Universal |
| Stratton et al[a] [102] | Disclosure decision aid tool | Australia (107) | - READY web-based disclosure decision aid tool<br>- 2-week access; 7 modules, requiring approximately 60 minutes to complete (53) | Information provided about disclosure on leading NGO[l] website (54) | Tertiary |
| Weber et al[a] [103] | Behavior change | Germany, England, and Northern Ireland (532) | - 4-week smartphone app “Kelaa Mental Resilience App”<br>- 28 sessions daily with 2 key modules<br>- Tracks mood and health and provides feedback; 6 to 7 daily sessions, each approximately 2 to 4 minutes (210) | WLC (322) | Universal |
| Deady et al[a] [104] | Behavior activation and mindfulness | Australia; male-dominated industries (2257) | - HeadGear, a smartphone app<br>- 30-day “challenge” daily (5-10 minutes per day)<br>- Risk calculator with personalized feedback, mood tracker, a toolbox of skills, and support service helplines (1131) | Active control same intervention but without risk calculator and mood tracker (1144) | Universal |

[a] Randomized controlled trial.

[b] CBT: cognitive behavioral therapy.

[c] WLC: wait-list control.

[d] Guided intervention.

[e] Controlled trial.

[f] ACT: acceptance and commitment therapy.

[g] MBSR: mindfulness-based stress reduction.

[h] NHS: National Health Service.

[i] PST: problem-solving therapy.

[j] MBRT: mindfulness-based resilience training.

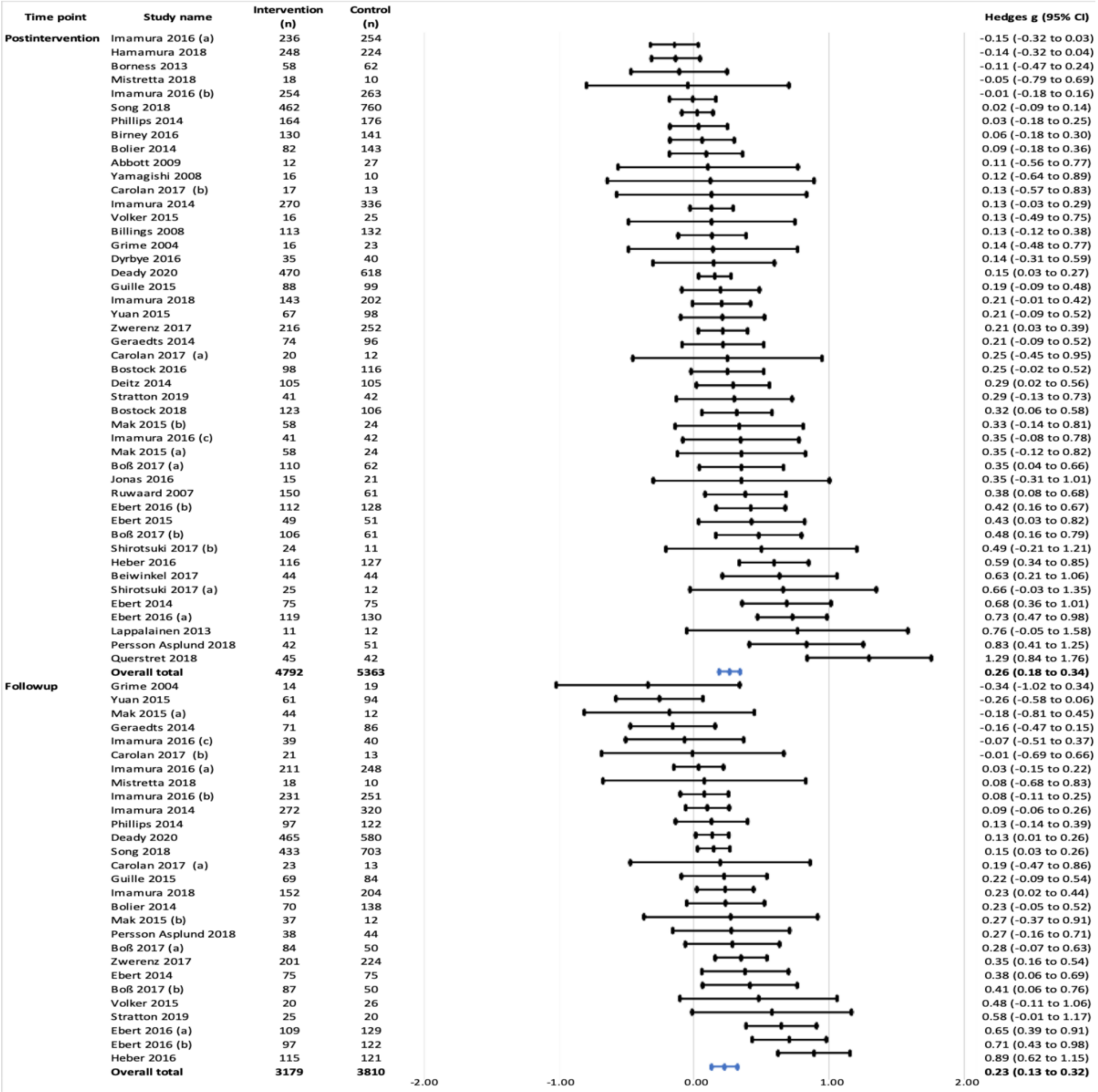
[k] MBCT: mindfulness-based cognitive therapy.

[l] NGO: nongovernmental organization.

### *Efficacy for Anxiety*

Overall, the 29 trials that measured anxiety outcomes showed a significant, small positive effect at the postintervention time point (Hedges $g$=0.26, 95% CI 0.13-0.39; $P$<.001), with high heterogeneity ($I^2$=77.49%; $\tau^2$=0.08). Approximately 21% (16/75) of studies reported follow-up outcomes, showing a small positive effect (Hedges $g$=0.32, 95% CI 0.15-0.50; $P$<.001), with similarly large heterogeneity ($I^2$=76.55%; $\tau^2$=0.08; Figure 3).

**Figure 3.** Effects on anxiety symptoms at postintervention and follow-up time points.



| Time point | Study name | Intervention (n) | Control (n) | Hedges g (95% CI) |
|---|---|---|---|---|
| **Postintervention** | Imamura 2016 (a) | 236 | 254 | -0.15 (-0.32 to 0.03) |
| | Hamamura 2018 | 248 | 224 | -0.14 (-0.32 to 0.04) |
| | Borness 2013 | 58 | 62 | -0.11 (-0.47 to 0.24) |
| | Mistretta 2018 | 18 | 10 | -0.05 (-0.79 to 0.69) |
| | Imamura 2016 (b) | 254 | 263 | -0.01 (-0.18 to 0.16) |
| | Song 2018 | 462 | 760 | 0.02 (-0.09 to 0.14) |
| | Phillips 2014 | 164 | 176 | 0.03 (-0.18 to 0.25) |
| | Birney 2016 | 130 | 141 | 0.06 (-0.18 to 0.30) |
| | Bolier 2014 | 82 | 143 | 0.09 (-0.18 to 0.36) |
| | Abbott 2009 | 12 | 27 | 0.11 (-0.56 to 0.77) |
| | Yamagishi 2008 | 16 | 10 | 0.12 (-0.64 to 0.89) |
| | Carolan 2017 (b) | 17 | 13 | 0.13 (-0.57 to 0.83) |
| | Imamura 2014 | 270 | 336 | 0.13 (-0.03 to 0.29) |
| | Volker 2015 | 16 | 25 | 0.13 (-0.49 to 0.75) |
| | Billings 2008 | 113 | 132 | 0.13 (-0.12 to 0.38) |
| | Grime 2004 | 16 | 23 | 0.14 (-0.48 to 0.77) |
| | Dyrbye 2016 | 35 | 40 | 0.14 (-0.31 to 0.59) |
| | Deady 2020 | 470 | 618 | 0.15 (0.03 to 0.27) |
| | Guille 2015 | 88 | 99 | 0.19 (-0.09 to 0.48) |
| | Imamura 2018 | 143 | 202 | 0.21 (-0.01 to 0.42) |
| | Yuan 2015 | 67 | 98 | 0.21 (-0.09 to 0.52) |
| | Zwerenz 2017 | 216 | 252 | 0.21 (0.03 to 0.39) |
| | Geraedts 2014 | 74 | 96 | 0.21 (-0.09 to 0.52) |
| | Carolan 2017 (a) | 20 | 12 | 0.25 (-0.45 to 0.95) |
| | Bostock 2016 | 98 | 116 | 0.25 (-0.02 to 0.52) |
| | Deitz 2014 | 105 | 105 | 0.29 (0.02 to 0.56) |
| | Stratton 2019 | 41 | 42 | 0.29 (-0.13 to 0.73) |
| | Bostock 2018 | 123 | 106 | 0.32 (0.06 to 0.58) |
| | Mak 2015 (b) | 58 | 24 | 0.33 (-0.14 to 0.81) |
| | Imamura 2016 (c) | 41 | 42 | 0.35 (-0.08 to 0.78) |
| | Mak 2015 (a) | 58 | 24 | 0.35 (-0.12 to 0.82) |
| | Boß 2017 (a) | 110 | 62 | 0.35 (0.04 to 0.66) |
| | Jonas 2016 | 15 | 21 | 0.35 (-0.31 to 1.01) |
| | Ruwaard 2007 | 150 | 61 | 0.38 (0.08 to 0.68) |
| | Ebert 2016 (b) | 112 | 128 | 0.42 (0.16 to 0.67) |
| | Ebert 2015 | 49 | 51 | 0.43 (0.03 to 0.82) |
| | Boß 2017 (b) | 106 | 61 | 0.48 (0.16 to 0.79) |
| | Shirotsuki 2017 (b) | 24 | 11 | 0.49 (-0.21 to 1.21) |
| | Heber 2016 | 116 | 127 | 0.59 (0.34 to 0.85) |
| | Beiwinkel 2017 | 44 | 44 | 0.63 (0.21 to 1.06) |
| | Shirotsuki 2017 (a) | 25 | 12 | 0.66 (-0.03 to 1.35) |
| | Ebert 2014 | 75 | 75 | 0.68 (0.36 to 1.01) |
| | Ebert 2016 (a) | 119 | 130 | 0.73 (0.47 to 0.98) |
| | Lappalainen 2013 | 11 | 12 | 0.76 (-0.05 to 1.58) |
| | Persson Asplund 2018 | 42 | 51 | 0.83 (0.41 to 1.25) |
| | Querstret 2018 | 45 | 42 | 1.29 (0.84 to 1.76) |
| | **Overall total** | **4792** | **5363** | **0.26 (0.18 to 0.34)** |
| **Followup** | Grime 2004 | 14 | 19 | -0.34 (-1.02 to 0.34) |
| | Yuan 2015 | 61 | 94 | -0.26 (-0.58 to 0.06) |
| | Mak 2015 (a) | 44 | 12 | -0.18 (-0.81 to 0.45) |
| | Geraedts 2014 | 71 | 86 | -0.16 (-0.47 to 0.15) |
| | Imamura 2016 (c) | 39 | 40 | -0.07 (-0.51 to 0.37) |
| | Carolan 2017 (b) | 21 | 13 | -0.01 (-0.69 to 0.66) |
| | Imamura 2016 (a) | 211 | 248 | 0.03 (-0.15 to 0.22) |
| | Mistretta 2018 | 18 | 10 | 0.08 (-0.68 to 0.83) |
| | Imamura 2016 (b) | 231 | 251 | 0.08 (-0.11 to 0.25) |
| | Imamura 2014 | 272 | 320 | 0.09 (-0.06 to 0.26) |
| | Phillips 2014 | 97 | 122 | 0.13 (-0.14 to 0.39) |
| | Deady 2020 | 465 | 580 | 0.13 (0.01 to 0.26) |
| | Song 2018 | 433 | 703 | 0.15 (0.03 to 0.26) |
| | Carolan 2017 (a) | 23 | 13 | 0.19 (-0.47 to 0.86) |
| | Guille 2015 | 69 | 84 | 0.22 (-0.09 to 0.54) |
| | Imamura 2018 | 152 | 204 | 0.23 (0.02 to 0.44) |
| | Bolier 2014 | 70 | 138 | 0.23 (-0.05 to 0.52) |
| | Mak 2015 (b) | 37 | 12 | 0.27 (-0.37 to 0.91) |
| | Persson Asplund 2018 | 38 | 44 | 0.27 (-0.16 to 0.71) |
| | Boß 2017 (a) | 84 | 50 | 0.28 (-0.07 to 0.63) |
| | Zwerenz 2017 | 201 | 224 | 0.35 (0.16 to 0.54) |
| | Ebert 2014 | 75 | 75 | 0.38 (0.06 to 0.69) |
| | Boß 2017 (b) | 87 | 50 | 0.41 (0.06 to 0.76) |
| | Volker 2015 | 20 | 26 | 0.48 (-0.11 to 1.06) |
| | Stratton 2019 | 25 | 20 | 0.58 (-0.01 to 1.17) |
| | Ebert 2016 (a) | 109 | 129 | 0.65 (0.39 to 0.91) |
| | Ebert 2016 (b) | 97 | 122 | 0.71 (0.43 to 0.98) |
| | Heber 2016 | 115 | 121 | 0.89 (0.62 to 1.15) |
| | **Overall total** | **3179** | **3810** | **0.23 (0.13 to 0.32)** |

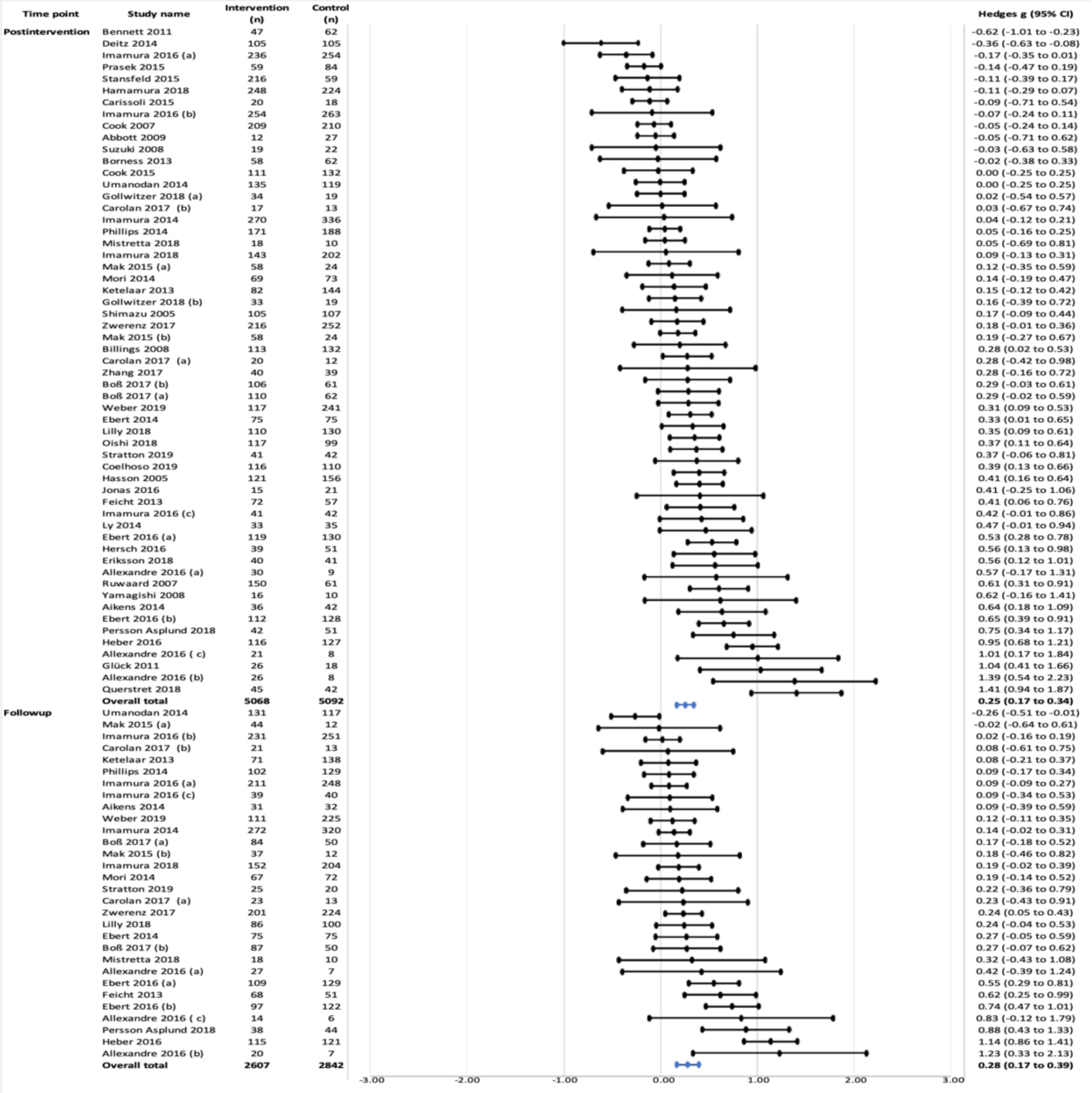
## *Efficacy for Depression*

An overall small positive effect on depression at the postintervention time point (Hedges *g*=0.26, 95% CI 0.19-0.34; *P*<.001) was estimated by combining 46 trials. Moderate heterogeneity was observed ($I^2$=66.96%; $\tau^2$=0.04). At the postintervention time point, one outlier was detected. By removing this outlier with a very large ES, a statistically significant but small magnitude positive effect remained (Hedges *g*=0.24, 95% CI 0.17-0.32; *P*<.001). Approximately 60% (45/75) of the studies reported follow-up outcomes. The follow-up effects on depression in 37% (28/75) of studies were very similar (Hedges *g*=0.23, 95% CI 0.13-0.32; *P*<.001), again with moderate heterogeneity ($I^2$=68.27%; $\tau^2$=0.03; Figure 2).

## *Efficacy for Stress*

Stress was the most common mental health outcome assessed in this study. In 76% (57/75) of studies, a small positive effect was found overall at the postintervention time point (Hedges *g*=0.25, 95% CI 0.17-0.34; *P*<.001). However, large heterogeneity was detected ($I^2$=76.11%; $\tau^2$=0.07). Overall, a significant but small positive effect was observed at follow-up in 40% (30/75) of studies (Hedges *g*=0.28, 95% CI 0.17-0.40; *P*<.001), with moderate heterogeneity detected ($I^2$=73.36%; $\tau^2$=0.06; Figure 4).

**Figure 4.** Effects on stress symptoms at postintervention and follow-up time points.



## *Moderating Factors for Outcome Efficacy*

### Anxiety

Mixed-effects meta-regression (Table 2) showed that for anxiety, the difference between approach efficacy was statistically significant ($Q$=21.72; $P$<.001; $R^2$=0.64). Stress management interventions (Hedges $g$=0.79, 95% CI 0.64-0.93; $P$<.001) and mindfulness (Hedges $g$=0.42, 95% CI 0.14-0.60; $P$<.001) interventions were more effective than CBT (Hedges $g$=0.11, 95% CI 0.04-0.19; $P$=.004) and *other* interventions (eg, cognitive training; Hedges $g$=0.13, 95% CI 0.01-0.25; $P$=.04). There were no significant between-group differences between the intervention approach, level of support, or type of control used.

**Table 2.** Meta-regression analysis for moderators of true effect on each mental health outcome at the postintervention time point.

| Intervention component and study design factor | Study, n (%) | Meta-analysis | | Heterogeneity | | | Meta-regression between-group tests | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Hedges *g* (95% CI) | *P* value | *Q* | *P* value | $I^2$ | *Q* | *P* value | $R^2$ |
| **Anxiety** | | | | | | | | | |
| **Type of intervention** | | | | | | | | | |
| CBT[a] | 14 (19) | 0.11 (0.04 to 0.19) | .004 | 14.41 | .35 | 9.78 | 21.72 | <.001 | 0.64 |
| Mindfulness | 5 (7) | 0.42 (0.14 to 0.6) | <.001 | 27.99 | <.001 | 85.71 | 21.72 | <.001 | 0.64 |
| Stress management | 4 (5) | 0.79 (0.64 to 0.93) | <.001 | 6.47 | .09 | 53.6 | 21.72 | <.001 | 0.64 |
| Other | 6 (8) | 0.13 (0.01 to 0.25) | .04 | 5.45 | .35 | 8.2 | 21.72 | <.001 | 0.64 |
| **Intervention approach** | | | | | | | | | |
| Indicated | 9 (12) | 0.44 (0.33 to 0.54) | <.001 | 60.13 | <.001 | 86.70 | 3.48 | .32 | 0.03 |
| Tailored | 3 (4) | 0.06 (−0.09 to 0.22) | .44 | 0.76 | .68 | 0.000 | 3.48 | .32 | 0.03 |
| Tertiary | 5 (7) | 0.14 (0.04 to 0.23) | .006 | 2.85 | .58 | 0.000 | 3.48 | .32 | 0.03 |
| Universal | 12 (16) | 0.29 (0.17 to 0.4) | <.001 | 38.03 | <.001 | 71.07 | 3.48 | .32 | 0.03 |
| **Person support** | | | | | | | | | |
| Yes | 14 (19) | 0.32 (0.24 to 0.4) | <.001 | 68.96 | <.001 | 81.15 | 1.02 | .31 | 0.01 |
| No | 15 (20) | 0.16 (0.07 to 0.24) | <.001 | 47.33 | <.001 | 70.42 | 1.02 | .31 | 0.01 |
| **Type of control** | | | | | | | | | |
| Active | 4 (5) | 0.26 (0.07 to 0.45) | .009 | 5.14 | .16 | 41.64 | 0.08 | .78 | 0 |
| WLC[b] | 25 (33) | 0.25 (0.19 to 0.31) | <.001 | 119.23 | <.001 | 79.87 | 0.08 | .78 | 0 |
| **Depression** | | | | | | | | | |
| **Type of intervention** | | | | | | | | | |
| CBT | 23 (31) | 0.11 (0.06 to 0.17) | <.001 | 44.12 | .003 | 50.13 | 20.32 | <.001 | 0.47 |
| Mindfulness | 5 (7) | 0.46 (0.28 to 0.64) | <.001 | 16.14 | .003 | 75.22 | 20.32 | <.001 | 0.47 |
| Stress management | 4 (5) | 0.61 (0.47 to 0.75) | <.001 | 4.03 | .26 | 25.52 | 20.32 | <.001 | 0.47 |
| Other | 14 (19) | 0.15 (0.09 to 0.21) | <.001 | 21.25 | .07 | 38.83 | 20.32 | <.001 | 0.47 |
| **Intervention approach** | | | | | | | | | |
| Indicated | 9 (12) | 0.32 (0.21 to 0.42) | <.001 | 45.95 | <.001 | 85.59 | 4.45 | .22 | 0.01 |
| Tailored | 6 (8) | 0.04 (−0.06 to 0.14) | .41 | 12.13 | .03 | 58.79 | 4.45 | .22 | 0.01 |
| Tertiary | 11 (15) | 0.2 (0.12 to 0.28) | <.001 | 18.98 | .04 | 47.31 | 4.45 | .22 | 0.01 |
| Universal | 20 (27) | 0.19 (0.13 to 0.25) | <.001 | 44.01 | .001 | 56.83 | 4.45 | .22 | 0.01 |
| **Person support** | | | | | | | | | |
| Yes | 20 (27) | 0.33 (0.27 to 0.4) | <.001 | 42.04 | .002 | 54.8 | 7.20 | .007 | 0.27 |
| No | 26 (35) | 0.11 (0.06 to 0.16) | <.001 | 65.58 | <.001 | 61.88 | 7.20 | .007 | 0.27 |
| **Type of control** | | | | | | | | | |
| Active | 9 (12) | 0.19 (0.11 to 0.28) | <.001 | 16.78 | .03 | 52.33 | 0.00 | .95 | 0.00 |
| WLC | 37 (49) | 0.18 (0.14 to 0.23) | <.001 | 119.35 | <.001 | 69.84 | 0.00 | .95 | 0.00 |
| **Stress** | | | | | | | | | |
| **Type of intervention** | | | | | | | | | |
| CBT | 19 (25) | 0.1 (0.04 to 0.16) | .001 | 47.89 | <.001 | 62.42 | 10.12 | .02 | 0.09 |
| Mindfulness | 15 (20) | 0.42 (0.31 to 0.53) | <.001 | 46.31 | <.001 | 69.77 | 10.12 | .02 | 0.09 |

| Intervention component and study design factor | Study, n (%) | Meta-analysis | | Heterogeneity | | | Meta-regression between-group tests | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Hedges *g* (95% CI) | *P* value | *Q* | *P* value | $I^2$ | *Q* | *P* value | $R^2$ |
| Stress management | 14 (19) | 0.28 (0.2 to 0.37) | .006 | 71.38 | <.001 | 88.79 | 10.12 | .02 | 0.09 |
| Other | 9 (12) | 0.12 (0.04 to 0.21) | <.001 | 37.74 | <.001 | 65.56 | 10.12 | .02 | 0.09 |
| **Intervention approach** | | | | | | | | | |
| Indicated | 9 (12) | 0.38 (0.28 to 0.49) | <.001 | 56.76 | <.001 | 85.91 | 3.26 | .35 | 0.00 |
| Tailored | 6 (8) | 0.07 (−0.03 to 0.16) | .16 | 27.10 | <.001 | 81.55 | 3.26 | .35 | 0.00 |
| Tertiary | 5 (7) | 0.12 (0.09 to 0.22) | .01 | 4.78 | .31 | 16.3 | 3.26 | .35 | 0.00 |
| Universal | 37 (49) | 0.18 (0.13 to 0.24) | <.001 | 125.06 | <.001 | 71.21 | 3.26 | .35 | 0.00 |
| **Person support** | | | | | | | | | |
| Yes | 22 (29) | 0.33 (0.26 to 0.39) | <.001 | 80.20 | <.001 | 73.82 | 6.79 | .009 | 0.15 |
| No | 35 (47) | 0.09 (0.04 to 0.14) | .001 | 121.76 | <.001 | 72.08 | 6.79 | .009 | 0.15 |
| **Type of control** | | | | | | | | | |
| Active | 7 (9) | 0.17 (0.05 to 0.29) | .006 | 18.38 | .005 | 67.36 | 0.03 | .87 | 0.00 |
| WLC | 50 (67) | 0.18 (0.14 to 0.23) | <.001 | 215.97 | <.001 | 77.31 | 0.03 | .87 | 0.00 |

[a] CBT: cognitive behavioral therapy.

[b] WLC: wait-list control.

### Depression

Similarly, significant differences were observed in the ESs for depression across the types of interventions (*Q*=21.72; *P*<.001; $R^2$=0.64). Again, stress management interventions (Hedges *g*=0.61, 95% CI 0.47-0.75; *P*<.001) and mindfulness (Hedges *g*=0.46, 95% CI 0.28-0.64; *P*<.001) interventions were more effective than CBT (Hedges *g*=0.11, 95% CI 0.06-0.17; *P*<.001) and *other* interventions (Hedges *g*=0.15, 95% CI 0.09-0.21; *P*<.001). Supported interventions had a higher ES (*g*=0.33, 95% CI 0.27-0.40; *P*<.001) than unsupported (Hedges *g*=0.11, 95% CI 0.06-0.16; *P*<.001) interventions (difference *Q*=7.20; *P*=.007; $R^2$=0.27). There were no significant differences between the intervention approach and the type of control used.

### Stress

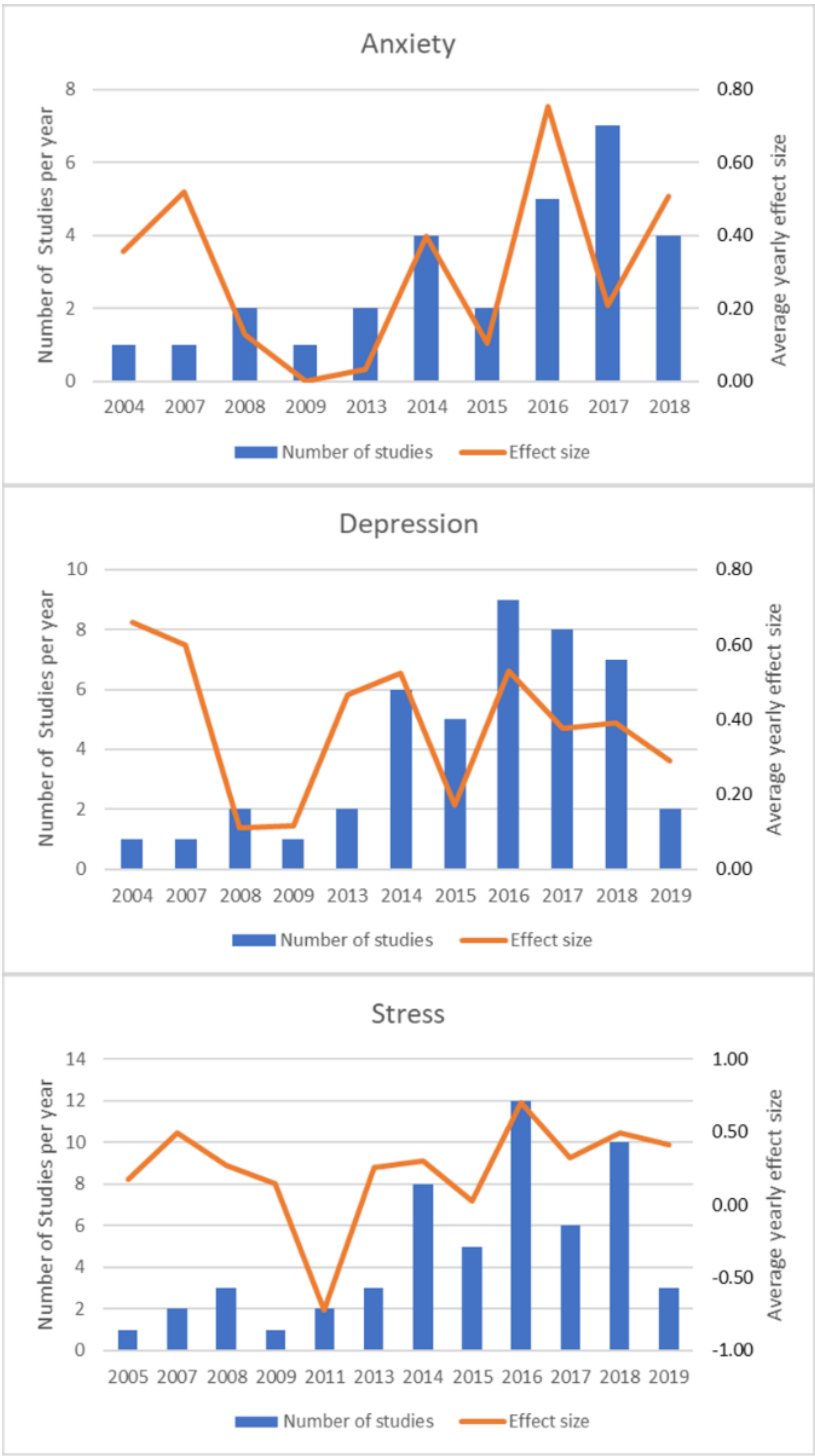
Differences were observed in the ESs across the type of intervention (*Q*=10.12; *P*=.02; $R^2$=0.09). Mindfulness interventions had the largest ES (Hedges *g*=0.42, 95% CI 0.31-0.53; *P*<.001), followed by stress management (Hedges *g*=0.28, 95% CI 0.20-0.37; *P*=.006), which was more effective than CBT (Hedges *g*=0.10, 95% CI 0.04-0.16; *P*=.001) and *other* interventions (Hedges *g*=0.12, 95% CI 0.04-0.21; *P*<.001). Supported interventions had a higher ES (Hedges *g*=0.33, 95% CI 0.26-0.39; *P*<.001) than unsupported (Hedges *g*=0.09, 95% CI 0.04-0.14; *P*=.001). The difference between these factors was statistically significant (*Q*=6.79; *P*=.009; $R^2$=0.15). For stress, there were no significant differences between the intervention approach and type of control used.

## Effectiveness of eHealth Interventions Over Time

### Overview

Over 5 times more studies evaluating eHealth interventions have been published since 2013 (n=64) than between 2004 and 2013 (n=11). However, the effectiveness of eHealth interventions did not seem to improve over time. The mean within-group ES reported in each year of the eHealth intervention arms in trials published since 2004 remained unchanged, with no significant SMD observed over time for anxiety ($F_{1,9}$=0.28; *P*=.61), depression ($F_{1,10}$=0.31; *P*=.59), or stress ($F_{1,11}$=0.75l; *P*=.41; Figure 5). There appears to have been a nadir in the effectiveness of studies published immediately after the financial crash and recession of 2008/2009.

**Figure 5.** Mean within-intervention group effect sizes reported in trials each year.



### Small Study Effect

A funnel plot for each mental health outcome is shown in Multimedia Appendix 6. No significant asymmetry was found for anxiety outcomes (n=29; Egger intercept 0.24; *P*=.80). However, significant asymmetry was observed for depression (n=46; Egger intercept 1.71; *P*=.001). After conducting a trim and fill analysis, 3% (2/75) of studies were imputed; the observed postintervention ES was adjusted to Hedges *g*=0.28 (95% CI 0.18-0.39). Similarly, significant asymmetry was observed in stress (n=57; Egger intercept 1.89; *P*=.004). After conducting a trim and fill analysis, 23% (17/75) of studies were imputed; the observed postintervention ES was adjusted to Hedges *g*=0.08 (95% CI 0.00-0.19), suggesting a greater effect reported in smaller studies.

## Discussion

### Principal Findings

This updated review aimed to synthesize the burgeoning literature on the efficacy of eHealth interventions for anxiety, depression, and stress outcomes in organizational settings and employee samples. We identified 52 new trial interventions published since the 23 identified in our prior review in 2017 [19], a tripling of the body of evidence. The systematic search identified 75 relevant trials for the meta-analysis, delivering eHealth interventions either on the web or via a smartphone, with a combined sample of 14,747 employees.

eHealth interventions reduced mental health symptoms immediately after use, with small positive effects observed in anxiety (Hedges $g$=0.26), depression (Hedges $g$=0.26), and stress (Hedges $g$=0.25), and data from trials with longer follow-up periods showed similar effects. These results are comparable with those of previous reviews, where small overall effects were found at the postintervention and follow-up time points [19,20]. These results imply that since 2017, the efficacy of eHealth interventions compared with control conditions reported in trials has remained unchanged, suggesting that the effectiveness of eHealth interventions does not seem to be improving over time. An analysis of the within-group effect observed in the intervention arms confirmed that since 2004, there has been no apparent systematic improvement in the effectiveness of these interventions. This is a surprising finding, given the enormous literature on methods to improve engagement with such interventions, greater penetration of technology to populations that do not access health care, and the increasing number of interventions delivered as apps, which reportedly improve access. All of these are commonly cited as factors favoring eHealth as a mode of intervention delivery or improving effectiveness.

The small but significant effect at follow-up suggests that eHealth interventions might have sustained positive effects on mental health. However, only half of the studies assessed follow-up outcomes, which may reflect a reporting bias. Supporting this, null, and in some cases, even negative, effects on mental health were observed in 27% (20/75) of the studies, and a greater proportion—60% (45/75)—did not report follow-up outcomes.

Less than one-fifth of all the trials used active controls. However, we found no significant differences in the efficacy of interventions tested against active versus passive controls overall or for any individual mental health outcome (Table 2). These findings do not confirm the *digital placebo effect*, which has been found in other eHealth reviews in clinical populations, where the ESs were lower in trials that used active controls than in those that used passive controls [24]. Given that most (47/75, 63%) of these trials were not conducted in clinically unwell populations but were delivered universally to employees, this contrary finding may have implications for the types of control recommended for different settings of eHealth trials and for future framework analysis or guidelines.

The primary moderator of efficacy appeared to be the content of the intervention, with stress management and mindfulness-based interventions being seemingly more efficacious than CBT-based interventions. This suggests that CBT, which was adapted from its intended use in a more personalized and clinical setting, may have less useful content, especially for use in organizations with universal delivery, than other approaches adapted for more universal or indicated tools such as mindfulness and stress management.

In our previous review, stronger effects were seen in eHealth interventions that were supported [19], and other studies have found increased effects when eHealth interventions were supported by trained mental health professionals [105]. The positive impacts of supported interventions may suggest that even in generally subclinical populations, a combination of eHealth and adjunctive support is the most effective in reducing mental health symptoms in employees.

Previous research has pointed to the importance of tailoring eHealth interventions to match individual user needs, as mental health symptoms differ from person to person [106]. However, this review did not support the notion that the tailoring of interventions provides any benefit to improving mental health. The limited evidence available suggests no greater efficacy when using tailored interventions than when using interventions delivered universally or for indicated or even unwell samples. In fact, 43% (3/7) of the tailored interventions showed potential for harm in at least one mental health outcome.

### Limitations

This study had some limitations. Significant heterogeneity was detected; however, this is not uncommon in this field or in meta-analytic research [24]. Differences across studies should be considered when interpreting these findings. It must also be acknowledged that some of the follow-up analyses were underpowered and that the findings should be interpreted tentatively. Furthermore, this review did not include any gray literature, and some authors did not provide data. Finally, the individual study pooled data calculated using the Comprehensive Meta-Analysis by Biostat Inc are not presented in this manuscript. Instead, we report the raw data of the pre-post means, SDs, and sample sizes in Table 1.

### Future Directions

As we were unable to detect that eHealth interventions are improving over time, to design effective eHealth interventions, a better understanding of the factors that may influence efficacy is required. This study did not consider engagement and adherence and only considered a small number of potential moderating factors for each intervention. A subsequent framework analysis model is being undertaken to establish the potential beneficial or harmful features of the different types of interventions. This is an important future direction and requires further in-depth analysis as eHealth interventions have the potential to offer a range of novel self-management tools for employees with clinical and subclinical mental health conditions.

Another recommendation is to develop a standard framework for eHealth interventions to best understand the features that have therapeutic benefits and those that may potentially cause harm. This will guide developers to ensure that eHealth interventions are designed in the most effective manner. A framework may address the significant heterogeneity within studies; if the development standards are regulated against this framework, the differences in the features of the interventions and how they are delivered may be reduced. This research suggests that those evaluating interventions or designing protocols should carefully consider the level of support provided when interpreting reported ESs.

## Conclusions

This review and meta-analysis confirms that eHealth interventions have a small but positive impact overall on reducing mental health symptoms in employees. There was significant heterogeneity between trials; however, overall, stress management and mindfulness interventions comprising in-person support appeared to be the most effective. Organizations should carefully consider the interventions delivered within the workplace; otherwise, they may not see long-term value in their return on investments. A substantial minority of intervention trials have demonstrated no efficacy, and a few may even be harmful. There is not enough evidence to make recommendations for the preferential use or development of therapeutically tailored interventions. Disappointingly, we found no evidence that despite the advancements in technology and the enormous resources in time, research, and finance devoted to this area for over a decade, better interventions are being produced. Hopefully, these small ESs do not represent the optimum outcome in organizational settings, and stakeholders and researchers should focus on improving effectiveness and efficacy or comparing and understanding the effects of current interventions. Despite the small effectiveness, this level of reduction could result in a large economic value in an organizational setting.

---

## Authors' Contributions

ES contributed to the conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing of the original draft, and review and editing of the manuscript. AL and IC contributed to data curation, methodology, and review and editing of the manuscript. IC contributed to data curation, methodology, and review and editing of the manuscript. HMG contributed to the data curation and review and editing of the manuscript. MA contributed to screening and study identification. JT contributed to the methodology, analysis plan, and review and editing of the manuscript. RAC and SMH contributed to the methodology and review and editing of the manuscript. NG contributed to data curation, investigation, methodology, supervision, and review and editing of the manuscript.

## Conflicts of Interest

NG has a patent PCT/AU2019/051186 pending. NG and ES own the IP of a web-based disclosure decision aid tool for employees.

Multimedia Appendix 1
Example of search terms.
[DOCX File , 38 KB - jmir_v24i9e37776_app1.docx ]

Multimedia Appendix 2
References of the included studies.
[DOCX File , 71 KB - jmir_v24i9e37776_app2.docx ]

Multimedia Appendix 3
Estimated risk of bias.
[PNG File , 278 KB - jmir_v24i9e37776_app3.png ]

Multimedia Appendix 4
Mental health measures.
[DOCX File , 28 KB - jmir_v24i9e37776_app4.docx ]

Multimedia Appendix 5
Detailed description of the identified studies and results.
[DOCX File , 101 KB - jmir_v24i9e37776_app5.docx ]

Multimedia Appendix 6
Overall studies small study effect.
[PNG File , 51 KB - jmir_v24i9e37776_app6.png ]

## References

1. Harvey SB, Modini M, Joyce S, Milligan-Saville JS, Tan L, Mykletun A, et al. Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. Occup Environ Med 2017 Mar;74(4):301-310. [doi: 10.1136/oemed-2016-104015] [Medline: 28108676]
2. Glozier N. Mental ill health and fitness for work. Occup Environ Med 2002 Oct;59(10):714-720 [FREE Full text] [doi: 10.1136/oem.59.10.714] [Medline: 12356935]
3. Harvey SB, Henderson M, Lelliott P, Hotopf M. Mental health and employment: much work still to be done. Br J Psychiatry 2009 Mar;194(3):201-203. [doi: 10.1192/bjp.bp.108.055111] [Medline: 19252144]
4. Leka S, Jain A, Iavicoli S, Vartia M, Ertel M. The role of policy for the management of psychosocial risks at the workplace in the European Union. Saf Sci 2011 Apr;49(4):558-564. [doi: 10.1016/j.ssci.2010.02.002]
5. Stevenson D, Farmer P. Thriving at Work: a review of mental health and employers. United Kingdom Government. 2017 Oct 26. URL: https://www.gov.uk/government/publications/thriving-at-work-a-review-of-mental-health-and-employers [accessed 2022-10-10]
6. Knudsen AK, Harvey SB, Mykletun A, Øverland S. Common mental disorders and long-term sickness absence in a general working population. The Hordaland Health Study. Acta Psychiatr Scand 2013 Apr;127(4):287-297. [doi: 10.1111/j.1600-0447.2012.01902.x] [Medline: 22775341]
7. Knudsen AK, Øverland S, Aakvaag HF, Harvey SB, Hotopf M, Mykletun A. Common mental disorders and disability pension award: seven year follow-up of the HUSK study. J Psychosom Res 2010 Jul;69(1):59-67. [doi: 10.1016/j.jpsychores.2010.03.007] [Medline: 20630264]
8. Harvey SB, Joyce S, Tan L, Johnson A, Nguyen H, Modini M, et al. Developing a mentally healthy workplace: a review of the literature – a report for the National Mental Health Commission and the Mentally Healthy Workplace Alliance. University of New South Wales. 2014 Nov. URL: https://www.headsup.org.au/docs/default-source/resources/developing-a-mentally-healthy-workplace_final-november-2014.pdf?sfvrsn=8 [accessed 2022-10-10]
9. Henderson M, Harvey SB, Overland S, Mykletun A, Hotopf M. Work and common psychiatric disorders. J R Soc Med 2011 May;104(5):198-207 [FREE Full text] [doi: 10.1258/jrsm.2011.100231] [Medline: 21558098]
10. Ebert DD, Kählke F, Buntrock C, Berking M, Smit F, Heber E, et al. A health economic outcome evaluation of an Internet-based mobile-supported stress management intervention for employees. Scand J Work Environ Health 2018 Mar 01;44(2):171-182 [FREE Full text] [doi: 10.5271/sjweh.3691] [Medline: 29144535]
11. Can Tech Help Manage Mental Health in the Workplace? CorporateWellnessMagazine. 2018. URL: https://www.corporatewellnessmagazine.com/article/tech-manage-mental-health-workplace [accessed 2020-08-04]
12. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
13. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
14. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression?-A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. BMC Psychiatry 2013 Apr 15;13:113 [FREE Full text] [doi: 10.1186/1471-244X-13-113] [Medline: 23587347]
15. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
16. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
17. McIntyre RS, Florea I, Tonnoir B, Loft H, Lam RW, Christensen MC. Efficacy of vortioxetine on cognitive functioning in working patients with major depressive disorder. J Clin Psychiatry 2017 Jan;78(1):115-121 [FREE Full text] [doi: 10.4088/JCP.16m10744] [Medline: 27780334]
18. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
19. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017 Dec 21;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
20. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of Web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]

21. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 01;45(6):560-576 [FREE Full text] [doi: 10.5271/sjweh.3839] [Medline: 31184758]
22. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
23. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
24. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. 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]
26. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. London, UK: The Cochrane Collaboration; Jan 01, 2013.
27. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Hedges L, Olkin I. Statistical Methods for Meta-Analysis. New York, NY, USA: Academic Press; 1985.
29. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons; 2009.
30. 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]
31. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
32. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 2010 Apr;1(2):97-111. [doi: 10.1002/jrsm.12] [Medline: 26061376]
33. Rothstein HR, Sutton AJ, Borenstein M. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. Chichester, UK: John Wiley & Sons; 2005.
34. 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]
35. Griffiths KM, Bennett K, Walker J, Goldsmid S, Bennett A. Effectiveness of MH-Guru, a brief online mental health program for the workplace: a randomised controlled trial. Internet Interv 2016 Nov;6:29-39 [FREE Full text] [doi: 10.1016/j.invent.2016.09.004] [Medline: 30135812]
36. Guille C, Zhao Z, Krystal J, Nichols B, Brady K, Sen S. Web-based cognitive behavioral therapy intervention for the prevention of suicidal ideation in medical interns: a randomized clinical trial. JAMA Psychiatry 2015 Dec;72(12):1192-1198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1880] [Medline: 26535958]
37. Imamura K, Furukawa TA, Matsuyama Y, Shimazu A, Kuribayashi K, Kasai K, et al. Differences in the effect of Internet-based cognitive behavioral therapy for improving nonclinical depressive symptoms among workers by time preference: randomized controlled trial. J Med Internet Res 2018 Aug 10;20(8):e10231 [FREE Full text] [doi: 10.2196/10231] [Medline: 30097419]
38. Ketelaar SM, Nieuwenhuijsen K, Gärtner FR, Bolier L, Smeets O, Sluiter JK. Effect of an E-mental health approach to workers' health surveillance versus control group on work functioning of hospital employees: a cluster-RCT. PLoS One 2013 Sep 12;8(9):e72546 [FREE Full text] [doi: 10.1371/journal.pone.0072546] [Medline: 24069148]
39. Volker D, Zijlstra-Vlasveld MC, Anema JR, Beekman AT, Brouwers EP, Emons WH, et al. Effectiveness of a blended Web-based intervention on return to work for sick-listed employees with common mental disorders: results of a cluster randomized controlled trial. J Med Internet Res 2015 May 13;17(5):e116 [FREE Full text] [doi: 10.2196/jmir.4097] [Medline: 25972279]
40. Cavanagh K, Churchard A, O'Hanlon P, Mundy T, Votolato P, Jones F, et al. A randomised controlled trial of a brief online mindfulness-based intervention in a non-clinical population: replication and extension. Mindfulness (N Y) 2018;9(4):1191-1205 [FREE Full text] [doi: 10.1007/s12671-017-0856-1] [Medline: 30100934]
41. Villani D, Grassi A, Cognetta C, Cipresso P, Toniolo D, Riva G. The effects of a mobile stress management protocol on nurses working with cancer patients: a preliminary controlled study. Stud Health Technol Inform 2012;173:524-528. [Medline: 22357050]
42. Geraedts AS, Kleiboer AM, Twisk J, Wiezer NM, van Mechelen W, Cuijpers P. Long-term results of a Web-based guided self-help intervention for employees with depressive symptoms: randomized controlled trial. J Med Internet Res 2014 Jul 09;16(7):e168 [FREE Full text] [doi: 10.2196/jmir.3539] [Medline: 25008127]

43. Geraedts AS, Kleiboer AM, Wiezer NM, van Mechelen W, Cuijpers P. Short-term effects of a Web-based guided self-help intervention for employees with depressive symptoms: randomized controlled trial. J Med Internet Res 2014 May 06;16(5):e121 [FREE Full text] [doi: 10.2196/jmir.3185] [Medline: 24800966]
44. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Effects of an Internet-based cognitive behavioral therapy (iCBT) program in Manga format on improving subthreshold depressive symptoms among healthy workers: a randomized controlled trial. PLoS One 2014 May 20;9(5):e97167 [FREE Full text] [doi: 10.1371/journal.pone.0097167] [Medline: 24844530]
45. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Does Internet-based cognitive behavioral therapy (iCBT) prevent major depressive episode for workers? A 12-month follow-up of a randomized controlled trial. Psychol Med 2015 Jul;45(9):1907-1917. [doi: 10.1017/S0033291714003006] [Medline: 25562115]
46. Eisen KP, Allen GJ, Bollash M, Pescatello LS. Stress management in the workplace: a comparison of a computer-based and an in-person stress-management intervention. Comput Human Behav 2008 Mar;24(2):486-496. [doi: 10.1016/j.chb.2007.02.003]
47. Wolever RQ, Bobinet KJ, McCabe K, Mackenzie ER, Fekete E, Kusnick CA, et al. Effective and viable mind-body stress reduction in the workplace: a randomized controlled trial. J Occup Health Psychol 2012 Apr;17(2):246-258. [doi: 10.1037/a0027278] [Medline: 22352291]
48. Boß L, Lehr D, Schaub MP, Paz Castro R, Riper H, Berking M, et al. Efficacy of a Web-based intervention with and without guidance for employees with risky drinking: results of a three-arm randomized controlled trial. Addiction 2018 Apr;113(4):635-646 [FREE Full text] [doi: 10.1111/add.14085] [Medline: 29105879]
49. Carolan S, Harris PR, Greenwood K, Cavanagh K. Increasing engagement with an occupational digital stress management program through the use of an online facilitated discussion group: results of a pilot randomised controlled trial. Internet Interv 2017 Dec;10:1-11 [FREE Full text] [doi: 10.1016/j.invent.2017.08.001] [Medline: 30135747]
50. Gollwitzer PM, Mayer D, Frick C, Oettingen G. Promoting the self-regulation of stress in health care providers: an Internet-based intervention. Front Psychol 2018 Jun 15;9:838 [FREE Full text] [doi: 10.3389/fpsyg.2018.00838] [Medline: 29962979]
51. Mak WW, Chan AT, Cheung EY, Lin CL, Ngai KC. Enhancing Web-based mindfulness training for mental health promotion with the health action process approach: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e8 [FREE Full text] [doi: 10.2196/jmir.3746] [Medline: 25599904]
52. Shirotsuki K, Nonaka Y, Abe K, Adachi SI, Adachi S, Kuboki T, et al. The effect for Japanese workers of a self-help computerized cognitive behaviour therapy program with a supplement soft drink. Biopsychosoc Med 2017 Sep 19;11:23 [FREE Full text] [doi: 10.1186/s13030-017-0109-5] [Medline: 28932258]
53. Allexandre D, Bernstein AM, Walker E, Hunter J, Roizen MF, Morledge TJ. A Web-based mindfulness stress management program in a corporate call center: a randomized clinical trial to evaluate the added benefit of onsite group support. J Occup Environ Med 2016 Mar;58(3):254-264 [FREE Full text] [doi: 10.1097/JOM.0000000000000680] [Medline: 26949875]
54. Grime PR. Computerized cognitive behavioural therapy at work: a randomized controlled trial in employees with recent stress-related absenteeism. Occup Med (Lond) 2004 Aug;54(5):353-359. [doi: 10.1093/occmed/kqh077] [Medline: 15289593]
55. Hasson D, Anderberg UM, Theorell T, Arnetz BB. Psychophysiological effects of a Web-based stress management system: a prospective, randomized controlled intervention study of IT and media workers [ISRCTN54254861]. BMC Public Health 2005 Jul 25;5:78 [FREE Full text] [doi: 10.1186/1471-2458-5-78] [Medline: 16042796]
56. Shimazu A, Kawakami N, Irimajiri H, Sakamoto M, Amano S. Effects of Web-based psychoeducation on self-efficacy, problem solving behavior, stress responses and job satisfaction among workers: a controlled clinical trial. J Occup Health 2005 Sep;47(5):405-413 [FREE Full text] [doi: 10.1539/joh.47.405] [Medline: 16230834]
57. Cook RF, Billings DW, Hersch RK, Back AS, Hendrickson A. A field test of a Web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. J Med Internet Res 2007 Jun 19;9(2):e17 [FREE Full text] [doi: 10.2196/jmir.9.2.e17] [Medline: 17581811]
58. Ruwaard J, Lange A, Bouwman M, Broeksteeg J, Schrieken B. E-mailed standardized cognitive behavioural treatment of work-related stress: a randomized controlled trial. Cogn Behav Ther 2007;36(3):179-192. [doi: 10.1080/16506070701381863] [Medline: 17852171]
59. Billings DW, Cook RF, Hendrickson A, Dove DC. A Web-based approach to managing stress and mood disorders in the workforce. J Occup Environ Med 2008 Aug;50(8):960-968 [FREE Full text] [doi: 10.1097/JOM.0b013e31816c435b] [Medline: 18695455]
60. Suzuki E, Tsuchiya M, Hirokawa K, Taniguchi T, Mitsuhashi T, Kawakami N. Evaluation of an Internet-based self-help program for better quality of sleep among Japanese workers: a randomized controlled trial. J Occup Health 2008 Oct 20;50(5):387-399 [FREE Full text] [doi: 10.1539/joh.l7154] [Medline: 18716392]
61. Yamagishi M, Kobayashi T, Nakamura Y. Effects of web-based career identity training for stress management among Japanese nurses: a randomized control trial. J Occup Health 2008;50(2):191-193 [FREE Full text] [doi: 10.1539/joh.l7086] [Medline: 18403870]
62. Abbott JA, Klein B, Hamilton C, Rosenthal A. The impact of online resilience training for sales managers on wellbeing and performance. E J Appl Psychol 2009 Jun 15;5(1):89-95. [doi: 10.7790/ejap.v5i1.145]

63. Bennett JB, Broome KM, Schwab-Pilley A, Gilmore P. A Web-based approach to address cardiovascular risks in managers: results of a randomized trial. J Occup Environ Med 2011 Aug;53(8):911-918 [FREE Full text] [doi: 10.1097/JOM.0b013e3182258bd8] [Medline: 21785368]
64. Glück TM, Maercker A. A randomized controlled pilot study of a brief Web-based mindfulness training. BMC Psychiatry 2011 Nov 08;11:175 [FREE Full text] [doi: 10.1186/1471-244X-11-175] [Medline: 22067058]
65. Borness C, Proudfoot J, Crawford J, Valenzuela M. Putting brain training to the test in the workplace: a randomized, blinded, multisite, active-controlled trial. PLoS One 2013;8(3):e59982 [FREE Full text] [doi: 10.1371/journal.pone.0059982] [Medline: 23555855]
66. Feicht T, Wittmann M, Jose G, Mock A, von Hirschhausen E, Esch T. Evaluation of a seven-week Web-based happiness training to improve psychological well-being, reduce stress, and enhance mindfulness and flourishing: a randomized controlled occupational health study. Evid Based Complement Alternat Med 2013;2013:676953 [FREE Full text] [doi: 10.1155/2013/676953] [Medline: 24489588]
67. Lappalainen P, Kaipainen K, Lappalainen R, Hoffrén H, Myllymäki T, Kinnunen ML, et al. Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. JMIR Res Protoc 2013 Jan 09;2(1):e1 [FREE Full text] [doi: 10.2196/resprot.2389] [Medline: 23611946]
68. Villani D, Grassi A, Cognetta C, Toniolo D, Cipresso P, Riva G. Self-help stress management training through mobile phones: an experience with oncology nurses. Psychol Serv 2013 Aug;10(3):315-322. [doi: 10.1037/a0026459] [Medline: 23937091]
69. Bolier L, Ketelaar SM, Nieuwenhuijsen K, Smeets O, Gärtner FR, Sluiter JK. Workplace mental health promotion online to enhance well-being of nurses and allied health professionals: a cluster-randomized controlled trial. Internet Interv 2014 Oct;1(4):196-204. [doi: 10.1016/j.invent.2014.10.002]
70. Deitz D, Cook RF, Hersch RK, Leaf S. Heart healthy online: an innovative approach to risk reduction in the workplace. J Occup Environ Med 2014 May;56(5):547-553 [FREE Full text] [doi: 10.1097/JOM.0000000000000148] [Medline: 24806568]
71. Ly KH, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. Internet Interv 2014 Jul;1(3):95-101. [doi: 10.1016/j.invent.2014.06.003]
72. Mori M, Tajima M, Kimura R, Sasaki N, Somemura H, Ito Y, et al. A Web-based training program using cognitive behavioral therapy to alleviate psychological distress among employees: randomized controlled pilot trial. JMIR Res Protoc 2014 Dec 02;3(4):e70 [FREE Full text] [doi: 10.2196/resprot.3629] [Medline: 25470499]
73. Phillips R, Schneider J, Molosankwe I, Leese M, Foroushani PS, Grime P, et al. Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention. Psychol Med 2014 Mar;44(4):741-752 [FREE Full text] [doi: 10.1017/S0033291713001323] [Medline: 23795621]
74. Umanodan R, Shimazu A, Minami M, Kawakami N. Effects of computer-based stress management training on psychological well-being and work performance in japanese employees: a cluster randomized controlled trial. Ind Health 2014;52(6):480-491 [FREE Full text] [doi: 10.2486/indhealth.2013-0209] [Medline: 25055847]
75. Carissoli C, Villani D, Riva G. Does a meditation protocol supported by a mobile application help people reduce stress? Suggestions from a controlled pragmatic trial. Cyberpsychol Behav Soc Netw 2015 Jan;18(1):46-53. [doi: 10.1089/cyber.2014.0062] [Medline: 25584730]
76. Cook RF, Hersch RK, Schlossberg D, Leaf SL. A Web-based health promotion program for older workers: randomized controlled trial. J Med Internet Res 2015 Mar 25;17(3):e82 [FREE Full text] [doi: 10.2196/jmir.3399] [Medline: 25830503]
77. Ebert DD, Berking M, Thiart H, Riper H, Laferton JA, Cuijpers P, et al. Restoring depleted resources: efficacy and mechanisms of change of an Internet-based unguided recovery training for better sleep and psychological detachment from work. Health Psychol 2015 Dec;34S:1240-1251. [doi: 10.1037/hea0000277] [Medline: 26651465]
78. Prasek A. Randomized Controlled Trial to Evaluate a Self-Guided, Web-Based Mindfulness Program for Stress Reduction and Wellbeing Promotion. University of Minnesota. 2015 Aug. URL: https://core.ac.uk/download/pdf/76357635.pdf [accessed 2022-08-29]
79. Stansfeld SA, Kerry S, Chandola T, Russell J, Berney L, Hounsome N, et al. Pilot study of a cluster randomised trial of a guided e-learning health promotion intervention for managers based on management standards for the improvement of employee well-being and reduction of sickness absence: GEM Study. BMJ Open 2015 Oct 26;5(10):e007981 [FREE Full text] [doi: 10.1136/bmjopen-2015-007981] [Medline: 26503383]
80. Yuan Q. Evaluating the Effectiveness of A Psychological Capital Development Program on Mental Health, Engagement and Work Performance. The Chinese University of Hong Kong. 2015. URL: https://www.proquest.com/docview/1846509447?pq-origsite=gscholar&fromopenview=true [accessed 2021-10-10]
81. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile Web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]

82. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
83. Dyrbye LN, West CP, Richards ML, Ross HJ, Satele D, Shanafelt TD. A randomized, controlled study of an online intervention to promote job satisfaction and well-being among physicians. Burnout Res 2016 Sep;3(3):69-75. [doi: 10.1016/j.burn.2016.06.002]
84. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided Internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
85. Ebert DD, Lehr D, Boß L, Riper H, Cuijpers P, Andersson G, et al. Efficacy of an Internet-based problem-solving training for teachers: results of a randomized controlled trial. Scand J Work Environ Health 2014 Nov;40(6):582-596 [FREE Full text] [doi: 10.5271/sjweh.3449] [Medline: 25121986]
86. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
87. Hersch RK, Cook RF, Deitz DK, Kaplan S, Hughes D, Friesen MA, et al. Reducing nurses' stress: a randomized controlled trial of a web-based stress management program for nurses. Appl Nurs Res 2016 Nov;32:18-25 [FREE Full text] [doi: 10.1016/j.apnr.2016.04.003] [Medline: 27969025]
88. Jonas B, Leuschner F, Tossmann P. Efficacy of an Internet-based intervention for burnout: a randomized controlled trial in the German working population. Anxiety Stress Coping 2017 Mar;30(2):133-144. [doi: 10.1080/10615806.2016.1233324] [Medline: 27602992]
89. Beiwinkel T, Eißing T, Telle N, Siegmund-Schultze E, Rössler W. Effectiveness of a Web-based intervention in reducing depression and sickness absence: randomized controlled trial. J Med Internet Res 2017 Jun 15;19(6):e213 [FREE Full text] [doi: 10.2196/jmir.6546] [Medline: 28619701]
90. Zhang H, Jiang Y, Nguyen HD, Poo DC, Wang W. The effect of a smartphone-based coronary heart disease prevention (SBCHDP) programme on awareness and knowledge of CHD, stress, and cardiac-related lifestyle behaviours among the working population in Singapore: a pilot randomised controlled trial. Health Qual Life Outcomes 2017 Mar 14;15(1):49 [FREE Full text] [doi: 10.1186/s12955-017-0623-y] [Medline: 28288636]
91. Zwerenz R, Becker J, Gerzymisch K, Siepmann M, Holme M, Kiwus U, et al. Evaluation of a transdiagnostic psychodynamic online intervention to support return to work: a randomized controlled trial. PLoS One 2017 May 8;12(5):e0176513 [FREE Full text] [doi: 10.1371/journal.pone.0176513] [Medline: 28481893]
92. Bostock S, Luik AI, Espie CA. Sleep and productivity benefits of digital cognitive behavioral therapy for insomnia: a randomized controlled trial conducted in the workplace environment. J Occup Environ Med 2016 Jul;58(7):683-689. [doi: 10.1097/JOM.0000000000000778] [Medline: 27257747]
93. Eriksson T, Germundsjö L, Åström E, Rönnlund M. Mindful self-compassion training reduces stress and burnout symptoms among practicing psychologists: a randomized controlled trial of a brief Web-based intervention. Front Psychol 2018 Nov 27;9:2340 [FREE Full text] [doi: 10.3389/fpsyg.2018.02340] [Medline: 30538656]
94. Hamamura T, Suganuma S, Ueda M, Mearns J, Shimoyama H. Standalone effects of a cognitive behavioral intervention using a mobile phone app on psychological distress and alcohol consumption among Japanese workers: pilot nonrandomized controlled trial. JMIR Ment Health 2018 Mar 22;5(1):e24 [FREE Full text] [doi: 10.2196/mental.8984] [Medline: 29567634]
95. Lilly M, Calhoun R, Painter I, Beaton R, Stangenes S, Revere D, et al. Destress 9-1-1-an online mindfulness-based intervention in reducing stress among emergency medical dispatchers: a randomised controlled trial. Occup Environ Med 2019 Oct;76(10):705-711. [doi: 10.1136/oemed-2018-105598] [Medline: 31138676]
96. Mistretta EG, Davis MC, Temkit M, Lorenz C, Darby B, Stonnington CM. Resilience training for work-related stress among health care workers: results of a randomized clinical trial comparing in-person and smartphone-delivered interventions. J Occup Environ Med 2018 Jun;60(6):559-568. [doi: 10.1097/JOM.0000000000001285] [Medline: 29370014]
97. Oishi S, Takizawa T, Kamata N, Miyaji S, Tanaka K, Miyaoka H. Web-based training program using cognitive behavioral therapy to enhance cognitive flexibility and alleviate psychological distress among schoolteachers: pilot randomized controlled trial. JMIR Res Protoc 2018 Jan 26;7(1):e32 [FREE Full text] [doi: 10.2196/resprot.8541] [Medline: 29374006]
98. Persson Asplund R, Dagöö J, Fjellström I, Niemi L, Hansson K, Zeraati F, et al. Internet-based stress management for distressed managers: results from a randomised controlled trial. Occup Environ Med 2018 Feb;75(2):105-113 [FREE Full text] [doi: 10.1136/oemed-2017-104458] [Medline: 28855344]
99. Querstret D, Cropley M, Fife-Schaw C. The effects of an online mindfulness intervention on perceived stress, depression and anxiety in a non-clinical sample: a randomised waitlist control trial. Mindfulness (N Y) 2018;9(6):1825-1836 [FREE Full text] [doi: 10.1007/s12671-018-0925-0] [Medline: 30524514]
100. Song M, Kanaoka H. Effectiveness of mobile application for menstrual management of working women in Japan: randomized controlled trial and medical economic evaluation. J Med Econ 2018 Nov;21(11):1131-1138 [FREE Full text] [doi: 10.1080/13696998.2018.1515082] [Medline: 30130990]

101. Coelhoso CC, Tobo PR, Lacerda SS, Lima AH, Barrichello CR, Amaro E, et al. A new mental health mobile app for well-being and stress reduction in working women: randomized controlled trial. J Med Internet Res 2019 Nov 07;21(11):e14269 [FREE Full text] [doi: 10.2196/14269] [Medline: 31697244]
102. Stratton E, Choi I, Calvo R, Hickie I, Henderson C, Harvey SB, et al. Web-based decision aid tool for disclosure of a mental health condition in the workplace: a randomised controlled trial. Occup Environ Med 2019 Sep;76(9):595-602. [doi: 10.1136/oemed-2019-105726] [Medline: 31413183]
103. Weber S, Lorenz C, Hemmings N. Improving stress and positive mental health at work via an app-based intervention: a large-scale multi-center randomized control trial. Front Psychol 2019 Dec 6;10:2745 [FREE Full text] [doi: 10.3389/fpsyg.2019.02745] [Medline: 31866915]
104. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
105. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
106. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]

---

## Abbreviations

**CBT:** cognitive behavioral therapy
**ES:** effect size
**PICOS:** Patient/Population, Intervention, Comparison, and Outcomes
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PROSPERO:** International Prospective Register of Systematic Reviews
**RCT:** randomized controlled trial
**SMD:** standardized mean difference

---

*Edited by R Kukafka; submitted 07.03.22; peer-reviewed by E Phillips, L Shen, V Nagy; comments to author 18.05.22; revised version received 26.05.22; accepted 29.07.22; published 27.09.22.*

*Please cite as:*
*Stratton E, Lampit A, Choi I, Malmberg Gavelin H, Aji M, Taylor J, Calvo RA, Harvey SB, Glozier N*
*Trends in Effectiveness of Organizational eHealth Interventions in Addressing Employee Mental Health: Systematic Review and Meta-analysis*
*J Med Internet Res 2022;24(9):e37776*
*URL: https://www.jmir.org/2022/9/e37776*
*doi:10.2196/37776*
*PMID:36166285*

Review

# Digital Interventions for Psychological Well-being in University Students: Systematic Review and Meta-analysis

Madeleine Ferrari[1], PhD; Stephanie Allan[2], MA; Chelsea Arnold[3], PhD; Dina Eleftheriadis[3], BA; Mario Alvarez-Jimenez[4,5], PhD; Andrew Gumley[2], PhD; John F Gleeson[3], PhD

[1]Healthy Brain and Mind Research Centre, Australian Catholic University, Sydney, Australia
[2]School of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom
[3]Healthy Brain and Mind Research Centre, Australian Catholic University, Melbourne, Australia
[4]Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia
[5]Orygen, Melbourne, Australia

**Corresponding Author:**
Madeleine Ferrari, PhD
Healthy Brain and Mind Research Centre
Australian Catholic University
25a Barker Road, Strathfield campus
Sydney, 2135
Australia
Phone: 61 0297014726
Email: madeleine.ferrari@acu.edu.au

## Abstract

**Background:** Life at university provides important opportunities for personal growth; however, this developmental phase also coincides with the peak period of risk for the onset of mental health disorders. In addition, specific university lifestyle factors, including impaired sleep and academic and financial stress, are known to exacerbate psychological distress in students. As a result, university students have been identified as a vulnerable population who often experience significant barriers to accessing psychological treatment. Digital psychological interventions are emerging as a promising solution for this population, but their effectiveness remains unclear.

**Objective:** This systematic review and meta-analysis aimed to assess digital interventions targeting psychological well-being among university students.

**Methods:** Database searches were conducted on December 2, 2021, via Embase, MEDLINE, PsycINFO, and Web of Science.

**Results:** A total of 13 eligible studies were identified, 10 (77%) of which were included in the meta-analysis. Mean pre-post effect sizes indicated that such interventions led to small and significant improvement in psychological well-being (Hedges $g$=0.32, 95% CI 0.23-0.4; $P$<.001). These effects remained, albeit smaller, when studies that included a wait-list control group were excluded (Hedges $g$=0.22, 95% CI 0.08-0.35; $P$=.002). An analysis of acceptance and commitment therapy approaches revealed small and significant effects ($k$=6; Hedges $g$=0.35, 95% CI 0.25-0.45; $P$<.001).

**Conclusions:** Digital psychological interventions hold considerable promise for university students, although features that optimize service delivery and outcomes require further assessment.

**Trial Registration:** PROSPERO CRD42020196654; https:/www.crd.york.ac.uk/prospero/display_record.php?RecordID=196654

***(J Med Internet Res 2022;24(9):e39686)*** doi:10.2196/39686

**KEYWORDS**

psychological well-being; mental health; university students; higher education students; college students; digital intervention; web-based intervention; mobile phone; meta-analysis; systematic review

# Introduction

## The Changing Landscape of Mental Well-being in University Students

Life at university provides important opportunities for personal growth during emerging adulthood by fostering autonomy, increasing social connectedness, and expanding intellectual horizons [1]. However, this developmental phase of emerging adulthood also coincides with the peak period of risk for onset of mental health disorders [2,3], and specific university lifestyle factors, including impaired sleep [4] and academic and financial stress, are known to exacerbate psychological distress in students [5,6].

Such risks are likely to contribute to the findings of the World Mental Health Surveys—International College Student Project by the World Health Organization, which indicates that between 20% and 31% of students experience a diagnoseable mental disorder [7,8]. Mental health symptoms at university entry are known to persist for at least the first year of study [6]. Furthermore, students with mental health disorders have been shown to eventually drop out from their studies at 2.5 times the rate of matched controls [9]. As the prevalence of mental health problems in young people has been steadily increasing [10], university counseling services have been unable to keep up with the growing demand; only 16% of students surveyed in the World Health Organization surveys with mental disorders reported receiving minimally adequate treatment [7].

## Reforming University Counseling and Mental Health Services

There have been recent calls for reform of university mental health services to address the increased prevalence of mental health problems among higher education students [11]. Duffy et al [11] proposed an integrated model of university-based mental health care that embeds multidisciplinary mental health teams within campus health services to provide timely care and to facilitate the transition from and to surrounding services. A significant challenge is that university students have a very broad spectrum of mental health needs, potentially ranging from thwarted personal growth in a narrow domain at one end of the spectrum (eg, suboptimal study habits) to enduring complex psychiatric disorders at the other end (eg, recurrent severe major depression combined with substance misuse). Intermediate needs may include mild and transient mental health symptoms (eg, self-limiting depressed mood) and incipient mental health episodes. There is also evidence that students with serious mental health problems have delayed access to mental health services, as evidenced by longer durations of untreated symptoms of psychosis [12]. The clinical staging model of mental health disorders by McGorryet al [13], which maps mental health interventions against commensurate levels of severity of mental health problems, was incorporated into the model by Duffy et al [11] as a guiding heuristic for responding to this complex spectrum. Specifically, the staging model highlights that the students presenting to university services range from those at early stages of risk with nonspecific mental health symptoms that do not meet criteria for a categorical diagnosis but who require monitoring and support to those at stage 3 and beyond who are recovering from a full threshold mental health disorder. This diversity of presenting problems poses a challenge in determining the effectiveness of mental health interventions for university students.

In the context of this diversity, psychological well-being (PWB) has emerged as an important construct with strong developmental face validity which should be considered when evaluating the effectiveness of university-based mental health interventions. PWB has been defined in terms of specific components of personal growth (including purpose in life, mastery, and self-acceptance) [14] and by the extent to which psychological needs, namely, autonomy, competence, and connectedness, have been met or thwarted (eg, by mental health symptoms) [15]. PWB also offers the advantage of being strongly correlated with psychopathology in youth but is also a distinct construct that may provide a basis for the prevention of mental health problems [16]. In addition, PWB has been identified as having strong transdiagnostic utility in predicting broad psychological outcomes across diverse populations [17,18]. Therefore, it is not surprising that PWB has emerged as a high priority for investigation and intervention in university students [19].

## The Role of Digital Technology in Mental Health Reform

A psychological intervention that seeks to improve mental health outcomes is considered digital when technology is used in its delivery, including the internet, mobile phones, computers, or other electronic devices. Digital interventions for mental health problems of varying severity have rapidly emerged as an innovation that promises improved access, acceptability, scalability, and cost-effectiveness compared with traditional face-to-face services alone [20]. Digital interventions for university students may circumvent students' concerns regarding stigma, time constraints, and lack of familiarity with health care systems [21]. In addition, high rates of access to smartphones and familiarity with mixed modes of learning mean that most university students are highly amenable to digital modes of health support [21]. Other global factors, such as the COVID-19 pandemic, have further accelerated the implementation of digital interventions as solutions for addressing limitations associated with traditional health care settings for young people [22].

Previous reviews related to this topic have reached contradictory conclusions in relation to PWB. Lattie et al [23] conducted a systematic review investigating the effectiveness, usability, acceptability, uptake, and adoption of digital mental health interventions for university students across a range of outcomes; however, the findings in relation to PWB were not reported separately from other outcomes. Inspection of their supplementary materials showed that across the 89 included publications, 8 studies specifically examined PWB as an outcome. One study measured PWB using a qualitative methodology and found little meaningful improvement, whereas the remaining studies consisted of randomized controlled trials (RCTs) and non-RCT and consistently reported significant improvement using validated measures (RCT: $k$=6; non-RCT: $k$=1). In contrast, a systematic review and meta-analysis of RCTs on the same topic and published in the same year reported

4 RCT studies that explicitly reported a measure of PWB (among other outcomes) when examining a web-based intervention for university students [24]. Although the review reported small and significant improvements in depression and anxiety symptoms, an analysis of the 4 studies that assessed PWB found nonsignificant effects for this outcome [24]. These conflicting results call for a closer examination of the effectiveness of digital interventions for PWB among university students.

There have been recent calls to reform mental health support for university students [11] and to develop stepped care models of psychological intervention to ease demands on existing mental health services [13]. As a result, it is necessary to review the evidence for the effectiveness and quality of digital interventions for PWB in university students. Therefore, the primary aim of this review was to examine the evidence for the effectiveness of digital mental health interventions for university students, specifically in relation to PWB. We also aimed to review the quality of this evidence, and given the diversity of this population, it is especially important to understand the range and severity of mental health symptoms reported in these studies. Our review question, in accordance with the PICO (population, intervention, control, and outcomes) framework [25], was whether web-based digital interventions, compared with active and passive control conditions, improve PWB in university student populations.

## Methods

### Search Strategy and Selection Criteria

The systematic review and meta-analysis were conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [25-27]. We conducted a review of trials, including pilot studies, non-RCTs, and RCTs, that evaluated any form of mental health or psychological health intervention targeting university students that was delivered through a web-based or eHealth medium, including mobile phone–based apps. Eligibility criteria included the assessment of a student population from a university or higher education institution (such as a college) of any age or nationality. Eligible studies were also required to be published in peer-reviewed journals in English language during and since 2000. Cohort, case-controlled, and cross-sectional studies were excluded because the study designs did not provide conclusions regarding the effectiveness of the interventions. For inclusion in this review, PWB needed to be assessed as a primary or secondary outcome using a validated measure. PWB is a multifaceted construct, and in this review, it was defined in line with the eudaemonic theories of well-being by Ryff [14] and Ryan et al [28]. Ryff [14] identified 6 core components of PWB that shape healthy development across the life span: self-acceptance, personal growth, purpose in life, environmental mastery, autonomy, and positive relations with others. The model of PWB by Ryff [14] provides a framework for meaning and purpose in life and has been operationalized using measures such as the Psychological Well-Being Scale and the Mental Health Continuum [29]. The self-determination theory by Ryan et al [28] postulates that motivation and wellness can be developed by meeting one's basic psychological needs: competence, relatedness, and autonomy [15]. The construct of PWB does not include symptoms of distress; thus, measures of distress and psychopathology, such as the depression, anxiety, and stress scale [30], were not included as primary outcome measures for the purpose of the review. The construct of PWB also does not include hedonism; thus, measures of happiness or life satisfaction, such as the Satisfaction with Life Scale [31], were also excluded from this review.

Searches were conducted on December 2, 2021, via Embase (Elsevier), MEDLINE (Ovid), PsycINFO (EBSCOhost), and Web of Science (Thomson Reuters). The search terms used were synonyms for *university student* and *digital intervention* (the full list of search terms and sample search syntax for each individual database is available in Multimedia Appendix 1). We contacted the corresponding authors via email to request further information where clarification of individual studies was required (eg, potential overlap in data across ≥2 published reports). In addition, reference lists of eligible studies and review articles were manually searched. The search was conducted by MF with consultation from a librarian. After deduplication in EndNote X8 (Clarivate Analytics), the completed search was imported into Covidence [32] for screening and data extraction. Abstract and full-text screening was conducted by CA, SA, and MF, with each paper being screened by at least two authors. The interrater reliability for independent full-text screening was reasonable ($k$=0.53; $P$=.76), and after conflicts were resolved by consultation between team members, there was a complete consensus. Data extraction was performed using a template developed by the authors in Excel and undertaken by CA, SA, and JG.

### Data Analysis

The extracted variables included publication characteristics (eg, authors, country, and year of publication), participant characteristics (eg, sample size; mean age; ethnicity; recruitment strategy, and depression, anxiety, and stress symptoms at baseline), study characteristics (eg, study design, primary aim of the study, and treatment conditions), PWB outcome measures (means and SDs at baseline, postintervention, and follow-up time points), and study findings. PWB was the primary outcome of interest. If studies reported multiple follow-up time points, data from the longest follow-up time point after the intervention were extracted.

Risk of bias (RoB) was assessed by CA and JG using the Cochrane collaboration RoB 2 tool [33] for RCTs, and the Downs and Black checklist was used for non-RCTs [34]. In addition, the Grading of Recommendations Assessment, Development, and Evaluation approach was used to evaluate the overall quality of the evidence [35].

We planned to perform a meta-analysis if 3 RCTs enabled the calculation of effect sizes in relation to PWB as a continuous variable. Meta-analysis was performed using Comprehensive Meta-Analysis (version 3), with random effects models used. A sensitivity analysis was performed to ascertain the effect of the wait-list versus active control groups. The $Q$ statistic was used to assess study heterogeneity [36]. We also calculated the

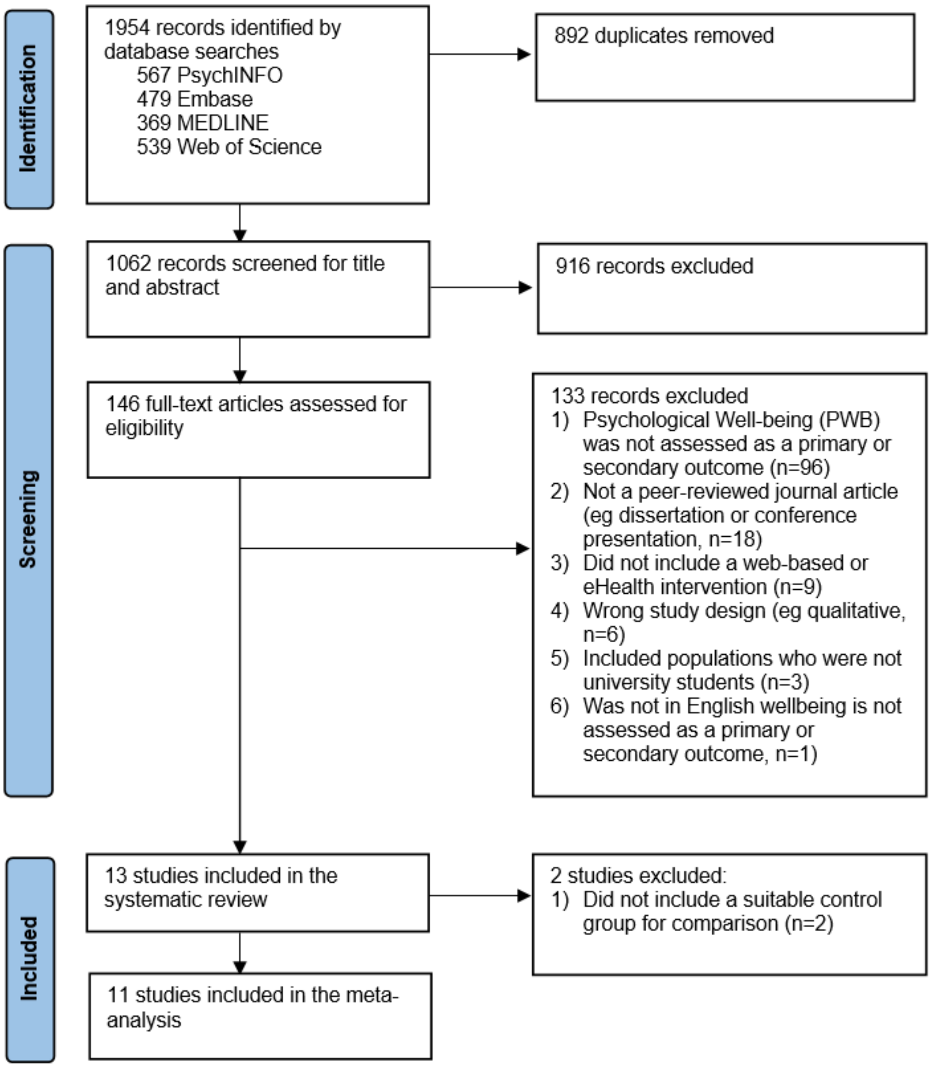
$I^2$ statistic to estimate the percentage of variance in the observed effects owing to the variance in the true effects. Heterogeneity can be considered low, moderate, substantial, or considerable, with $I^2$ values of 0% to 40%, 30% to 60%, 30% to 90%, and 75% to 100%, respectively [37]. The protocol was registered with PROSPERO before screening (registration number: CRD42020196654).

## Results

### Search and Selection

A total of 1954 references were imported into Covidence (Veritas Health Innovation) for initial abstract screening. After removing 892 duplicates, 1062 studies were screened against title and abstract, from which 916 studies were excluded. In all, 146 studies were assessed for full eligibility via full-text screening; the reasons for exclusion are mentioned in Figure 1. Finally, 13 studies met the inclusion criteria and were included in this review [38-50]. Of these 13 studies, 11 (85%) were eligible for inclusion in the meta-analysis [38-42,45,46,48-51].

**Figure 1.** Study selection.



### Study Characteristics

Table 1 summarizes the characteristics of the included studies. All 13 studies were published from 2016 onward, with 5 (38%) conducted in the United States [39-41,47,49]; 2 (15%) in Sweden [43,48] and Australia [42,44]; and 1 (8%) in Hong Kong [46], Finland [50], Ireland [45], and the United Kingdom [38]. The sample sizes across the 13 studies ranged from 23 [40] to 2110 [44]. Regarding study design, there were 46% (6/13) of RCTs with an active control [39,40,43,45,46,48], 38% (5/13) of RCTs with a wait-list control [38,41,42,49,52], 8% (1/13) were dismantling trials [41], and 8% (1/13) were pre-post trials [47]. The duration of the digital interventions varied, with one study reporting a single session [43], another study reporting 6 days of intervention [44], and the remaining studies (11/13, 85%) ranging from 3 to 10 weeks ($\text{mean}_{\text{duration}}$=5.5, SD 2.2 weeks; mode duration=4 weeks), as presented in Table 1 [35-42,45-47]. Across all studies, outcomes were measured at pre- and postintervention time points, and the duration of follow-up assessment intervals from baseline ranged from 7 days [46] to 12 months [50].

**Table 1.** Characteristics of studies included in the systematic review.

| Study and country | Primary aim | Population | | Study features | | Intervention | | Outcomes and effect size |
|---|---|---|---|---|---|---|---|---|
| | | N; recruitment; completers (%)[a] | Demographics | Study design | Measurement; time points; PWB[b] outcome measure | Intervention type; treatment target; delivery mode; duration | Comparison group | |
| Auyeung et al [46], 2019; Hong Kong | Examine the outcome efficacy of the Best Possible Self intervention in improving well-being | 139; university (web-based via university mass mail) and social media; 68.6% | Age: T[c]-22.94 years (SD 3.02; n=48), C[d]-22.7 years (SD 3.68; n=52); female: T-72.9% (n=48), C-73.1% (n=52); ethnicity: NR[e] | Pilot RCT[f] (T vs active C) | Pre, post; TFS[g] | Best Possible Self Positive Psychology intervention; well-being (increase flourishing and decrease depression); web-based mobile and desktop versions; 6 days | Active: participants wrote about details of 5 events from the past 24 hours | *Flourishing*[h]*:* time × condition effect favoring T. Partial $\eta^2$=0.12; *autonomy:* time x condition effect favoring T. Partial $\eta^2$=0.04; *relatedness:* no time x condition effect. ES[i] not reported; *conclusion:* intervention more effective at improving flourishing and autonomy compared with active control with moderate effect sizes. No improvements in competence or relatedness |
| Bendtsen et al [51], 2020; Sweden | To compare the positive psychology intervention with control | 654; 15 universities (digital advertising); 61.2% | Age: T-25 years (median age), C-26 years (median age control); female: T-79.6% (n=277), C-76.1% (n=233); ethnicity: NR | RCT (T C-TAU[j]) | Pre, 3 months after randomization; MHC-SF[k] | mHealth[l] positive psychology multicomponent program; positive mental health; via smartphone; 10 weeks | Active: web-based mental health information control group. Sent via SMS text message | *Psychological well-being:* group x time effect favoring T. IRR[m]=1.067 (95% CI 1.024 to 1.112); *conclusion:* psychological well-being significantly higher at 3 months in intervention group compared with control group with small effect size |
| Firestone et al [47], 2019; United States | Test the usability, acceptability, receptivity, and utility to promote valued living and psychological well-being of the LYV[n] program | 137; university (psychology classes); 84% | Age: 20.22 years (SD 4.35); female: 86.9%; White: 53.3% | Single group: post | Pre, post; SPWB[o]: autonomy subscale | Living Your Values: ACT[p]-based; valued living; web-based, self-guided values-focused program; single 60- to 90-minute session | None | No significant treatment effects (within-group pre-post) on any well-being subscales. ESs not reported; *conclusion:* no evidence that the intervention was effective at improving psychological well-being |

| Study and country | Primary aim | Population | | Study features | | Intervention | | Outcomes and effect size |
|---|---|---|---|---|---|---|---|---|
| | | N; recruitment; completers (%)[a] | Demographics | Study design | Measurement; time points; PWB[b] outcome measure | Intervention type; treatment target; delivery mode; duration | Comparison group | |
| Kvillemo et al [48], 2016; Sweden | Examine the feasibility, usability, acceptability, and outcomes of an 8-week internet-based mindfulness training program | 90; university; 39.1% | Age: T completers-29 years (range 18–45 years, n=18), T noncompleters-24 years (range 19-37, n=22); female: T completers-88.9% (n=18) T noncompleters-63.6% (n=14); ethnicity: NR | Pilot RCT (T vs active C) | Pre, post; SPWB: total score | Mindfulness training program; mindfulness; internet-based; 8 weeks | Active: internet-based expressive writing intervention. Participants could make contact with study coordinators via phone or email | *Psychological well-being:* no time x group interaction. T had a statistical increase in psychological well-being over time. Cohen *d*=0.2; *conclusion:* no evidence for group effect in relation to improving well-being. Treatment group showed statistical and small improvement in well-being from preintervention to postintervention |
| Levin et al [39], 2016; United States | Test the feasibility of a web-based ACT prototype prevention program called ACT-CL[q] | 234; university (digital advertising); 70% | Age: 21.61 years (SD 5.48); female: 76.9%; White: 6.2%, Asian: 9.3%, Black or African American: 3.5%, American Indian or Alaska native: 1.8%, native Hawaiian or other pacific islander: 1.3%, and other: 7.9% | Pilot RCT (T vs active C) | Pre, post, 1-month FU[r], 3-month FU; MHC-SF | ACT-CL; psychological symptoms, positive mental health, and psychological flexibility; web-based, 2 core multimedia sessions and supplementary emails, web-based resources, and SMS text messages; 3 weeks | Active: 2 session mental health education website (length matched to ACT-CL) basic educational information | *Psychological well-being:* no significant group x time effect. Effect size not reported; *Conclusion:* the effects of the ACT-CL program were largely equivalent to those of an education website, with a lower level of program engagement with ACT-CL |
| Levin et al [49], 2017; United States | Evaluate a transdiagnostic web-based self-help program that includes all ACT components and is designed to treat a wide range of problems | 79; university (2 cohorts; fall 2014 and spring 2015); 80% | Age: 20.51 years (SD 2.73); female: 66%; White: 88% | Pilot RCT (T v wait-list C) | Pre, post; MHC-SF | ACT; mental health problems; self-help website; 4 weeks | Wait-list | *Positive mental health:* time x condition effect favoring T Cohen *d*=0.58; *emotional well-being:* no significant time x condition interaction. ES not reported; *social well-being:* time x condition effect favoring T Cohen *d*=0.69; *Psychological well-being:* no significant time x condition effect. ES not reported; *Conclusion:* the intervention was not more effective than wait-list control for improving psychological well-being |

| Study and country | Primary aim | Population | | Study features | | Intervention | | Outcomes and effect size |
|---|---|---|---|---|---|---|---|---|
| | | N; recruitment; completers (%)[a] | Demographics | Study design | Measurement; time points; PWB[b] outcome measure | Intervention type; treatment target; delivery mode; duration | Comparison group | |
| Levin et al [40], 2020; United States | Evaluate the feasibility and acceptability of a popular mindfulness meditation app (SBT[s]) for students on a college counseling center wait-list | 23; university (counseling service); 60% | Age: 20.43 years (SD 2.46); female: 100%; non-Hispanic White: 87% | Pilot RCT (T v wait-list C) | Pre, during (2 weeks), Post (4 weeks after baseline); MHC-SF | SBT; mindfulness; mobile app; 4 weeks | Wait-list | *Psychological well-being:* statistical effects not calculated owing to small sample size. Pre-post T favored SBT for positive mental health: Hedges *g*=0.52 (95% CI−0.31 to 1.41); *Conclusion:* no evidence for statistical effect on well-being for intervention compared with wait-list control. Small to moderate effect size favoring intervention at postintervention |
| Levin et al [41], 2020; United States | Compare web-based versions of ACT targeting the open components, the engaged components, or full ACT, relative to a wait-list condition | 181; university; 88.9% | Age: 22.27 (SD 5.08 years); female-72.4%; White: 92.8% | Randomized dismantling trial with 4 conditions (full ACT vs active, open vs active, engaged vs wait-list control) | Pre, post, 4-week FU; MHC-SF | 12- session web-based ACT intervention (with differing components); acceptance, cognitive diffusion (open), values, committed action (engaged); via computer or smartphone; 6 weeks | 2 Active; open components (open); engaged components (engaged); and wait-list control | *Psychological well-being:* time x condition effect – full or engaged>wait-list (pre-post, pre-FU), wait-list did not differ from open, engaged>open (pre-post). Cohen *d* wait-list vs full=0.51, wait-list vs engaged=0.69, engaged vs open=0.56; *Conclusion:* “Engaged” and “full” intervention, but not “open” intervention more effective at improving positive mental health compared with wait-list with medium effect sizes |
| Noone et al [45], 2018; Ireland | Investigate if regular mindfulness meditation practice facilitates critical thinking through the enhancement of executive function | 91; university; 72.9% | Age: 20.92 years (SD 4.39); female: 76%; ethnicity: NR | RCT (T v active C) | Pre, post; WEMWBS[t] | Headspace mindfulness; mindfulness; mobile app; 6 weeks | Active: Sham mindfulness app | *Subjective well-being:* group assignment not associated with well-being at FU; b=2.01, 95% CI –0.63 to 4.65; *Conclusion:* no evidence that intervention more effective than active control for improving subjective well-being |

| Study and country | Primary aim | Population | | Study features | | Intervention | | Outcomes and effect size |
|---|---|---|---|---|---|---|---|---|
| | | N; recruitment; completers (%)[a] | Demographics | Study design | Measurement; time points; PWB[b] outcome measure | Intervention type; treatment target; delivery mode; duration | Comparison group | |
| Ponzo et al [38], 2020; United Kingdom | Test the efficacy of a 4-week intervention delivered via a mobile app and wearable device (BioBase program) in comparison with a wait-list control group | 262; university; 45.4% | Age: T at baseline- 19.9 years (SD 1.83, n=72), C at baseline- 19.84 (1.76, n=74); female: T at baseline- 62.5%, C at baseline- 63.5%; ethnicity: NR | RCT (T vs wait-list C) | Pre, 2 weeks, Post (4 weeks), 2-week FU (6 week); WEMWBS | BioBase: eclectic approach with mindfulness, biofeedback interventions, CBT[u], and behavioral activation theory; well-being; via smartphone and wearable device; 4 weeks | Wait-list | *Perceived well-being:* group x time effect (post and FU); pre-post (within T) Cohen *d*=0.65, pre-FU (within T) Cohen *d*=1.16; *Conclusion:* intervention effective at improving well-being compared with wait-list control with sustained (2 week) effects at medium-large effect sizes |
| Räsänen et al [50], 2016; Finland | Evaluate the efficacy of the web-based Student Compass program including 2 face-to-face meetings, tailored individual written feedback on the web, coping tools, and strategies | 68; university; 87.9% | Age: 24.29 years (SD 3.28); female: 85.3%; ethnicity: NR | RCT (T vs wait-list C) | Pre, post, 12-month FU (intervention only); MHC-SF | ACT intervention; stress, anxiety, depression; guided, blended web-based and face-to-face; 7 weeks | Wait-list | *Psychological well-being:* group x time effect favoring T; significant improvement over time from baseline to 12 months FU in treatment group. Between pre-post corrected Cohen *d*=0.46; within pre-post corrected Cohen *d*=0.61; within pre-12- month FU corrected Cohen *d*=0.65; *Conclusion:* treatment more effective than wait-list control at improving psychological well-being with medium effect sizes |
| Viskovich et al [44], 2018; Australia | Pilot-test a web-based ACT mental health promotion program called YOLO[v] for university students | 130; university; 40% | Age: 26.34 years (SD 7.96); female: 75.4%; White or Australian: 53.1% | Pilot RCT, 3 groups vary format: 1 (weekly flexible); 2 (full flexibility); 3 (sequential) | Pre, post; MHC-SF | YOLO ACT program; cognitive fusion, acceptance, mindfulness, values, and committed action; web-based; 4 weeks | 2 Active: full flexibility delivery and sequential delivery | *Psychological well-being:* T significantly improved from preintervention to postintervention on the primary outcome of well-being across 3 delivery formats, Cohen *d*=0.25; *Conclusion:* intervention associated with improved well-being from pre to post time points with small effect size |

| Study and country | Primary aim | Population | | Study features | | Intervention | | Outcomes and effect size |
|---|---|---|---|---|---|---|---|---|
| | | N; recruitment; completers (%)[a] | Demographics | Study design | Measurement; time points; PWB[b] outcome measure | Intervention type; treatment target; delivery mode; duration | Comparison group | |
| Viskovich et al [42], 2020; Australia | Evaluate the effectiveness of a web-based ACT mental health promotion intervention called YOLO in an RCT | 2110; university; 29.3% | Age: 26.85 years (SD 8.77); female: 67.8%; ethnicity: NR | RCT (T vs wait-list C) | Pre, post; 12-week FU; MHC-SF | YOLO ACT program; cognitive fusion, acceptance, mindfulness, values, and committed action; web-based consisting of 4 modules; 4 weeks | Wait-list | *Psychological well-being:* time x condition effect favoring T Cohen *d*=0.37; *Conclusion:* intervention more effective than wait-list control at improving well-being with small effect size |

[a] Completers (%): percentage of participants who completed the postintervention data collection by the number of participants who were randomized to the intervention group.

[b] PWB: psychological well-being.

[c] T: treatment group.

[d] C: control group.

[e] NR: not reported.

[f] RCT: randomized controlled trial.

[g] TFS: The Flourishing Scale.

[h] Key outcomes and study conclusions italicized.

[i] ES: effect size.

[j] C-TAU: control-treatment as usual.

[k] MHC-SF: Mental Health Continuum−Short Form [53].

[l] mHealth: mobile health.

[m] IRR: incidence rate ratio.

[n] LYV: living your values.

[o] SPWB: Ryff Scales of Psychological Well-being-42 item [54].

[p] ACT: acceptance and commitment therapy.

[q] ACT-CL: ACT on college life.

[r] FU: follow-up.

[s] SBT: Stop, Breathe & Think.

[t] WEMWBS: Warwick-Edinburgh Mental Well-being Scale [55].

[u] CBT: cognitive behavioral therapy.

[v] YOLO: You Only Live Once.

## Mental Ill-Health Symptoms

The predominant measure of PWB was the Mental Health Continuum-Short Form [56] used in 62% (8/13) of the studies [39-44,49,50], followed by the scales of PWB [54] reported in 15% (2/13) of the studies [47,48], the Warwick-Edinburgh Mental Well-being Scale [55] reported in 15% (2/13) of the studies [38,45], and the Flourishing Scale [57] reported in 8% (1/13) of the studies [46]. A measure of the severity of mental health symptoms was included in 85% (11/13) of the studies and excluded in the other 15% (2/13) [45,47]. The Depression Anxiety Stress Scales (DASS) [30,58] were used in 31% (4/13) of studies [38,39,42,44], the Counseling Center Assessment of Psychological Symptoms was used in 23% (3/13) of studies [40,41,49], the Center for Epidemiological Studies–Depression [59,60] was used in 15% (2/13) of studies [46,48], and the Perceived Stress Scale [61] and the Hospital Anxiety and Depression Scale (HADS) [62] were used in 8% (1/13) of studies [43,50].

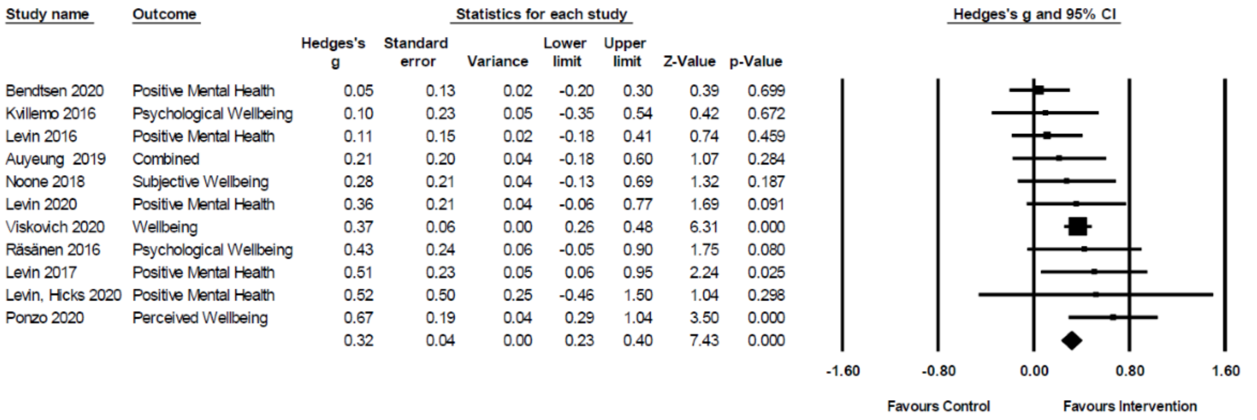
To better understand the prevalence and severity of mental ill-health symptoms in the general university student population, the DASS and HADS baseline scores were further analyzed using a software program that estimates the percentile scores and interval estimates for individual scores [63]. Of the studies that used these measures, some reported symptom cut-off scores as eligibility criteria [38,43] or targeted students who self-identified as distressed [50] or were actively seeking psychological treatment [41] and thus tended to report higher symptoms of mental ill-health. Studies using the DASS and HADS, which did not specify such eligibility criteria or targeted recruitment, may reflect the general mental health of university students. Levin et al [39] reported elevated depression (88th percentile, 95% CI 86-90) and anxiety (86th percentile, 95% CI 84-88) symptoms in university students compared with the general population. Viskovich et al [44] also reported elevated depression (86th percentile, 95% CI 84-88) and anxiety symptoms (90th percentile, 95% CI 88-92). At baseline, both the intervention and control groups reported by Viskovich et al [42] displayed elevated depression (intervention–93rd percentile, 95% CI 92-94; control 94th percentile, 95% CI 94-95), anxiety (intervention–94th percentile, 95% CI 93-95; control–94th percentile, 95% CI 93-95), and stress symptoms (intervention–96th percentile, 95% CI 95-97; control–97th percentile, 95% CI 96-98).

## Intervention Effectiveness

Of the 13 studies, 11 (85%) were eligible for inclusion in this meta-analysis [38-42,45,46,48-51]. Two corresponding authors were contacted and provided additional data needed for the meta-analysis, which were not reported in the published papers [39,51]. A study was excluded because the study design comprised a single-group pre-post comparisons [47]. In addition, a further study was excluded from the meta-analysis, as it reported on 3 intervention groups, each of which delivered identical content but in different formats (as planned, full flexibility and sequentially), and thus, it was not deemed to include an appropriate comparative group to address the core research question of this review [44].

Of the 11 studies eligible for inclusion in the meta-analysis [38-42,45,46,48-51], the aggregate effect of treatment on PWB, as displayed in Figure 2, was small and statistically significant compared with controls ($k$=11; n=2903; Hedges $g$=0.32, 95% CI 0.23-0.4; $P$<.001). There was no evidence of significant heterogeneity ($Q_{10}$=12.71; $P$=.24; $I^2$=21.34; $T^2$=0.01; $t$=0.08). Sensitivity analyses were performed to examine whether the treatment effect differed across studies with active (7/11, 64%) [38,39,45,46,48,50,51] and wait-list control (4/11, 36%) comparison groups [40-42,49]. When wait-list control studies were removed, the effect was reduced but remained statistically significant (Hedges $g$=0.22, 95% CI 0.08-0.35; $P$=.001).

**Figure 2.** Effect of digital psychological interventions on psychological well-being in university student populations [38-42,45,46,48-51].



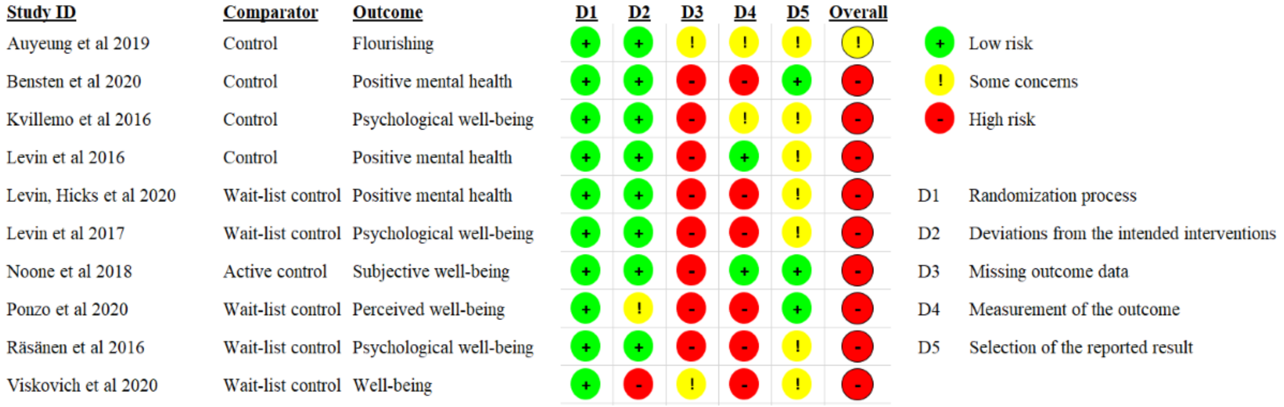
| Study name | Outcome | Hedges's g | Standard error | Variance | Lower limit | Upper limit | Z-Value | p-Value |
|---|---|---|---|---|---|---|---|---|
| Bendtsen 2020 | Positive Mental Health | 0.05 | 0.13 | 0.02 | -0.20 | 0.30 | 0.39 | 0.699 |
| Kvillemo 2016 | Psychological Wellbeing | 0.10 | 0.23 | 0.05 | -0.35 | 0.54 | 0.42 | 0.672 |
| Levin 2016 | Positive Mental Health | 0.11 | 0.15 | 0.02 | -0.18 | 0.41 | 0.74 | 0.459 |
| Auyeung 2019 | Combined | 0.21 | 0.20 | 0.04 | -0.18 | 0.60 | 1.07 | 0.284 |
| Noone 2018 | Subjective Wellbeing | 0.28 | 0.21 | 0.04 | -0.13 | 0.69 | 1.32 | 0.187 |
| Levin 2020 | Positive Mental Health | 0.36 | 0.21 | 0.04 | -0.06 | 0.77 | 1.69 | 0.091 |
| Viskovich 2020 | Wellbeing | 0.37 | 0.06 | 0.00 | 0.26 | 0.48 | 6.31 | 0.000 |
| Räsänen 2016 | Psychological Wellbeing | 0.43 | 0.24 | 0.06 | -0.05 | 0.90 | 1.75 | 0.080 |
| Levin 2017 | Positive Mental Health | 0.51 | 0.23 | 0.05 | 0.06 | 0.95 | 2.24 | 0.025 |
| Levin, Hicks 2020 | Positive Mental Health | 0.52 | 0.50 | 0.25 | -0.46 | 1.50 | 1.04 | 0.298 |
| Ponzo 2020 | Perceived Wellbeing | 0.67 | 0.19 | 0.04 | 0.29 | 1.04 | 3.50 | 0.000 |
| | | 0.32 | 0.04 | 0.00 | 0.23 | 0.40 | 7.43 | 0.000 |

All interventions were stand-alone, meaning they were offered to university students independent of university counseling services, and run by independent researchers or an independent youth mental health service such as Headspace [45]. Some studies used university services for recruitment, such as posting advertisements on existing student health care center websites [51] or waiting rooms [42], approaching students on wait-lists for counseling centers [40], or advertising through university student affairs [41]. Of the 13 studies included in the systematic review, acceptance and commitment therapy (ACT) was the most frequently adopted therapeutic orientation, present in 7 (54%) studies [39,41,42,44,47,49,50]. In addition, 23% (3/13) of studies adopted mindfulness interventions [40,45,48], 15% (2/13) used positive psychological interventions [43,46], and 8% (1/13) described an eclectic intervention [38]. A total of 46% (6/13) of interventions were web-based [39,42,44,47-49], 31% (4/13) were mobile apps [38,40,43,45], 15% (2/13) were available via mobile or desktop modes [41,46], and 8% (1/13) combined web-based and face-to-face modes of delivery [50]. Given the large number of ACT intervention studies, we conducted a subgroup analysis to evaluate the effectiveness of this therapeutic modality in relation to PWB. The meta-analytic effect of ACT-based treatments on PWB was small and statistically significant (Hedges $g$=0.35, 95% CI 0.25-0.45; $P$<.001).

## Risk of Bias

In relation to the study quality, based on study design, 77% (10/13) studies were rated according to the RoB-2 [33] and 23% (3/13) using the Downs and Black checklist [34]. As shown in Figure 3, overall, 8% (1/13) of studies were rated on the RoB-2 with “some concerns” of risk in relation to bias [46], and 69% (9/13) were rated as “high risk” of bias [38,39,41,42,45,48-51]. Missing outcome data were the most prevalent domain of concern, followed by the measurement of outcomes resulting from reliance on self-report measures. A total of 23% (3/13) of non-RCTs were rated using the Downs and Black checklist [34]. Each study was given a total score based on performance across 5 domains (reporting; external validity; internal validity-bias; internal validity-confounding; and power); each paper was scored (yes=1; no or unable to determine=0). Moreover, 15% (2/13) of studies were classified as overall fair quality [41,47] and 8% (1/13) as poor quality [44], according to classifications of the total score used in other reviews [64,65]. Common areas of weakness across 23% (3/13) of the studies assessed included a lack of reporting of adverse events, lack of attempt to blind participants or researchers, and not clearly reporting the intended analysis in methods, a priori. The overall estimation of the quality of evidence, based on the Grading of Recommendations Assessment, Development, and Evaluation assessment, was generally moderate (Multimedia Appendix 2).

**Figure 3.** Risk of bias assessment using the Cochrane risk of bias tool (version 2) [38-42,45,46,48-51].



| Study ID | Comparator | Outcome | D1 | D2 | D3 | D4 | D5 | Overall |
|---|---|---|---|---|---|---|---|---|
| Auyeung et al 2019 | Control | Flourishing | + | + | ! | ! | ! | ! |
| Bensten et al 2020 | Control | Positive mental health | + | + | - | - | + | - |
| Kvillemo et al 2016 | Control | Psychological well-being | + | + | - | ! | ! | - |
| Levin et al 2016 | Control | Positive mental health | + | + | - | + | ! | - |
| Levin, Hicks et al 2020 | Wait-list control | Positive mental health | + | + | - | - | ! | - |
| Levin et al 2017 | Wait-list control | Psychological well-being | + | + | - | - | ! | - |
| Noone et al 2018 | Active control | Subjective well-being | + | + | - | + | + | - |
| Ponzo et al 2020 | Wait-list control | Perceived well-being | + | ! | - | - | + | - |
| Räsänen et al 2016 | Wait-list control | Psychological well-being | + | + | - | - | ! | - |
| Viskovich et al 2020 | Wait-list control | Well-being | + | - | ! | - | ! | - |

+ Low risk
! Some concerns
- High risk

D1 Randomization process
D2 Deviations from the intended interventions
D3 Missing outcome data
D4 Measurement of the outcome
D5 Selection of the reported result

A further RoB is study retention, which refers to the number of participants who completed the research study compared with those who discontinued the study or did not complete data collection at ≥1 time points. The most consistent retention data that could be collated across studies were the comparison of the number of participants randomized to the intervention with the number of participants who completed data collection at the post–time point for the specified primary outcome measures. The proportion of participants who completed postintervention data collection varied from 29.3% to 88.9% across the included studies, with an average of 63.6% completers across all studies (Table 1). Few studies have reported reasons for participant drop out, and the details of such reporting varied widely. In addition, most studies did not explicitly report adherence to the prescribed treatment. A study reported that although the reason for failing to complete the intervention was not systematically assessed, 9 participants emailed the coordinator and provided a reason, including lack of time, technical problems with software, or changed personal circumstances [45]. In relation to publication bias, visual inspection of the funnel plot suggested a largely symmetrical distribution, indicating a high likelihood that this review has captured all relevant studies and presents a low risk of publication bias for PWB (Multimedia Appendix 3). The Egger test of the intercept was not significant (intercept –0.09, SE 0.566; *P*=.44). Using the trim-and-fill method by Duval and Tweedie [66,67], only minor changes in values were observed, further supporting a low risk of publication bias.

## Discussion

### Principal Findings

This study aimed to synthesize the published literature on the effectiveness of digital mental health interventions for university students, specifically in relation to PWB. Our systematic review and meta-analysis found small to moderate effects on improving PWB in university students following exposure to a digital intervention. Beneficial effects remained when studies with a wait-list control group were removed, indicating that the effects of digital interventions remained robust in comparison with active controls. It is notable that ACT was the most prevalent theoretical framework underlying psychological interventions in the included studies. A subgroup analysis of 7 ACT-based interventions showed significant and small to moderate improvements in PWB. ACT targets transdiagnostic processes, such as the identification of values [68], which provide developmentally salient therapeutic targets for university students, while also accommodating heterogeneity in clinical needs. ACT as a therapeutic modality also coherently aligns with PWB as a key outcome, in addition to symptom alleviation [69-71]. Given that past reviews have identified PWB and flourishing as important mental health outcomes [17,18], especially for university students [19], the findings of this review provide further support for the development of digital psychological interventions grounded in ACT.

Our review resolves prior conflicting findings on the effectiveness of digital interventions for improving PWB in university students [23,24] and extends this work by exclusively focusing on PWB outcomes. Our findings partially align with the systematic review by Lattie et al [23], which identified 8 studies examining PWB as an outcome from digital mental health interventions for university students. The authors reported that 1 non-RCT and 6 RCTs found significant improvements using validated measures, whereas 1 study using qualitative measures did not find meaningful improvement. Our findings are in contrast with the meta-analysis conducted by Harrer et al [24], which found nonsignificant effects of digital interventions for university students on PWB across 4 RCTs. The specific focus of our review on PWB provides support for the capacity of digital interventions to meaningfully improve this specific psychological construct for university students, although these improvements were modest. Our review also identified ACT-based approaches as the most common therapeutic model used in the included studies, which also resulted in small to modest effects on PWB.

The baseline mental health data from the included studies indicated the presence of severe symptoms of mental ill-health in university student populations. Other studies have previously found that 20% to 31% of students experience a diagnoseable mental disorder [7,8]. The studies included in this review measured symptomology, not diagnosable mental health disorders, and found that depression symptoms in university students were more severe than that of 86% to 94% of the general population [39,42,44], anxiety symptoms were more severe than that of 86% to 94% of the general population [39,42,44], and stress symptoms more severe than that of 95% to 96% [42] of the general population. It is promising that this review found small to moderate and significant improvements in PWB as a result of digital interventions, despite the severity of symptoms. The recent COVID-19 pandemic is likely to place additional stressors on university students and exacerbate symptoms [72].

The COVID-19 global pandemic has had, and will continue to have, far-reaching consequences on individual, societal, and global functioning [73]. Although higher education rapidly pivoted to web-based learning in Western countries to reduce face-to-face contact and slow the spread of the virus [74], all 13 studies included in this review were conducted before the pandemic. It was noted during the literature search that many web-based psychological interventions were conducted for university students in 2020 and 2021; however, these were often excluded from this review because PWB was not an outcome. We argue that PWB is of even greater importance during a pandemic. The results of this review support the capacity of digital psychological interventions to improve PWB for university students; however, overall effect sizes were modest. Future research could target innovation to enhance the effectiveness of web-based interventions. Such innovation may include developing a better understanding of the predisposing and precipitating triggers of poor PWB among university students. Further innovation may also include identifying the active components of ACT approaches that address university students’ needs and consider approaches to integrate digital interventions within existing university-based care models.

Recent calls for the reform of university mental health services [11] are warranted. Duffy et al [11] recently emphasized the importance of integrated models of university-based mental health care that met the spectrum of student psychological needs [13] and embedded multidisciplinary mental health teams within campus services. Such models would comprehensively provide preventive and urgent treatment to high-risk populations that is likely to result in both immediate relief and long-term improvement in mental health trajectories [11]. In this review, all included studies featured stand-alone interventions, meaning they were not integrated with existing counseling or mental health support services; however, some studies used such services for recruitment purposes. For example, Levin et al [40] approached students on wait-lists for university counseling centers and provided a digital intervention while waiting for face-to-face appointments, effectively filling a gap in clinical need. In addition, Räsänen et al [50] reported that their web-based ACT intervention was also available more broadly in a Swedish university to 15,000 enrolled students in 2 formats: a self-help offering and twice a year offering in a coach-supported form. Firestone et al [47] concluded that such digital interventions could be integrated with university orientation programs in the future.

## Strengths and Limitations

This review provides a stronger evidence base to support the recommendations [11,13] for integrating effective digital interventions with existing counseling and student support services on university campuses, particularly when such interventions are grounded in an ACT therapeutic modality. This conclusion aligns with existing research that young adults are likely to endorse blended models of psychological care [75]. Future interventions could also consider the timing of such interventions; for example, before exams or during orientation may be when mental health needs are particularly salient to students. In addition, only 1 study included a 12-month follow-up after the intervention [50], highlighting the need for future research to ascertain whether the treatment effects of digital interventions persist over time for university students.

There are several methodological concerns regarding the quality of the available studies. Study retention is a commonly recognized challenge for digital psychological interventions [76,77], with a meta-analysis reporting dropouts from such programs ranging from 2% to 83%, with a weighted average of 31% [78]. In this review, we calculated study retention based on the proportion of participants who completed the postintervention data compared with those randomized or allocated to the intervention group. The proportion of completers varied from 29.3% to 88.9% across the included studies, with an average of 63.6% completing the digital intervention, similar to completion rates reported in previous reviews [73]. Future studies on digital interventions for university students should systematically collect both treatment adherence and study compliance data, accompanied by explicit reasons for drop out or discontinuation of the treatment.

In addition, a group of researchers conducted 3 of the included studies [39,40,49], potentially resulting in an undue influence of a paradigm or researcher’s approach or style of intervention. Given that 2 of these studies presented some of the strongest effect sizes for PWB [40,49], it may be that the effect was the result of something specific to this group’s implementation of ACT. Alternatively, a significant, positive finding in 2 independent studies by the same group of researchers may also increase confidence that results are less likely to have been a chance finding, assuming that bias was carefully managed. Further examination of ACT-based digital psychological interventions across different university student populations would clarify the effectiveness of these approaches. As discussed, the fail-safe N and funnel plot analyses suggest that there is a low risk of publication bias. Overall, the quality of the studies was rated as moderate, with the most common methodological issues potentially causing bias including failure to report complete outcome data and issues with the measurement of PWB. Future research should be strengthened by reporting greater detail when describing intervention content to facilitate an understanding of the mechanisms of change, consideration of dosage effects, and an assessment of the acceptability of such interventions.

## Conclusions

Overall, the results of this systematic review and meta-analysis indicate that digital psychological interventions are a promising area of research and clinical intervention for enhancing PWB among university students. The most common therapeutic modality for digital interventions was ACT, which theoretically focuses on strengthening the individual's capacity to lead a rich and value-driven life, a goal which is consistent with PWB outcomes. The effect sizes demonstrated significant improvements in PWB for university students, albeit only modestly. The findings of this review encourage the further development of evidence-based digital interventions that target PWB in vulnerable populations such as university students.

## Acknowledgments

MAJ was supported by an investigator grant (APP1177235) from the National Health and Medical Research Council and a Dame Kate Campbell Fellowship from the University of Melbourne. MF was supported by an Open Access Publishing Support Grant from the Faculty of Health Sciences at the Australian Catholic University to cover the publication cost.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 23 KB - jmir_v24i9e39686_app1.docx ]

Multimedia Appendix 2
Grading of Recommendations Assessment, Development, and Evaluation quality assessment.
[DOCX File , 22 KB - jmir_v24i9e39686_app2.docx ]

Multimedia Appendix 3
Funnel plot of studies included in the random-effects meta-analysis of digital psychological interventions for psychological well-being in university students compared with control groups.
[DOCX File , 46 KB - jmir_v24i9e39686_app3.docx ]

## References

1. Arnett JJ. College students as emerging adults: the developmental implications of the college context. Emerg Adulthood 2015 May 24;4(3):219-222. [doi: 10.1177/2167696815587422]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022 Jan;27(1):281-295 [FREE Full text] [doi: 10.1038/s41380-021-01161-7] [Medline: 34079068]
4. Cheng SH, Shih CC, Lee IH, Hou YW, Chen KC, Chen KT, et al. A study on the sleep quality of incoming university students. Psychiatry Res 2012 May 30;197(3):270-274. [doi: 10.1016/j.psychres.2011.08.011] [Medline: 22342120]
5. Deasy C, Coughlan B, Pironom J, Jourdan D, Mannix-McNamara P. Psychological distress and coping amongst higher education students: a mixed method enquiry. PLoS One 2014 Dec 15;9(12):e115193 [FREE Full text] [doi: 10.1371/journal.pone.0115193] [Medline: 25506825]
6. Duffy A, Keown-Stoneman C, Goodday S, Horrocks J, Lowe M, King N, et al. Predictors of mental health and academic outcomes in first-year university students: identifying prevention and early-intervention targets. BJPsych Open 2020 May 08;6(3):e46 [FREE Full text] [doi: 10.1192/bjo.2020.24] [Medline: 32381150]
7. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
8. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]
9. Ishii T, Tachikawa H, Shiratori Y, Hori T, Aiba M, Kuga K, et al. What kinds of factors affect the academic outcomes of university students with mental disorders? A retrospective study based on medical records. Asian J Psychiatr 2018 Feb;32:67-72. [doi: 10.1016/j.ajp.2017.11.017] [Medline: 29216609]
10. Patalay P, Gage S. Changes in millennial adolescent mental health and health-related behaviours over 10 years: a population cohort comparison study. Int J Epidemiol 2019 Oct 01;48(5):1650-1664 [FREE Full text] [doi: 10.1093/ije/dyz006] [Medline: 30815691]

11. Duffy A, Saunders KE, Malhi GS, Patten S, Cipriani A, McNevin SH, et al. Mental health care for university students: a way forward? Lancet Psychiatry 2019 Nov;6(11):885-887. [doi: 10.1016/S2215-0366(19)30275-5] [Medline: 31324561]
12. Reichert A, Jacobs R. Socioeconomic inequalities in duration of untreated psychosis: evidence from administrative data in England. Psychol Med 2018 Apr;48(5):822-833. [doi: 10.1017/S0033291717002197] [Medline: 28805178]
13. McGorry PD, Nelson B, Goldstone S, Yung AR. Clinical staging: a heuristic and practical strategy for new research and better health and social outcomes for psychotic and related mood disorders. Can J Psychiatry 2010 Aug;55(8):486-497. [doi: 10.1177/070674371005500803] [Medline: 20723276]
14. Ryff CD, Keyes CL. The structure of psychological well-being revisited. J Pers Soc Psychol 1995 Oct;69(4):719-727. [doi: 10.1037//0022-3514.69.4.719] [Medline: 7473027]
15. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
16. Campbell S, Osborn TL. Adolescent psychopathology and psychological wellbeing: a network analysis approach. BMC Psychiatry 2021 Jul 03;21(1):333 [FREE Full text] [doi: 10.1186/s12888-021-03331-x] [Medline: 34217254]
17. Gleeson JF, Eleftheriadis D, Santesteban-Echarri O, Koval P, Bastian B, Penn DL, et al. Positive and meaningful lives: systematic review and meta-analysis of eudaimonic well-being in first-episode psychosis. Early Interv Psychiatry 2021 Oct;15(5):1072-1091. [doi: 10.1111/eip.13049] [Medline: 33037789]
18. Fava GA, Guidi J. The pursuit of euthymia. World Psychiatry 2020 Feb;19(1):40-50 [FREE Full text] [doi: 10.1002/wps.20698] [Medline: 31922678]
19. Dodd AL, Priestley M, Tyrrell K, Cygan S, Newell C, Byrom NC. University student well-being in the United Kingdom: a scoping review of its conceptualisation and measurement. J Ment Health 2021 Jun;30(3):375-387. [doi: 10.1080/09638237.2021.1875419] [Medline: 33567937]
20. Batra S, Baker RA, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl) 2017 Oct 4;10:237-251 [FREE Full text] [doi: 10.2147/MDER.S144158] [Medline: 29042823]
21. Montagni I, Cariou T, Feuillet T, Langlois E, Tzourio C. Exploring digital health use and opinions of university students: field survey study. JMIR Mhealth Uhealth 2018 Mar 15;6(3):e65 [FREE Full text] [doi: 10.2196/mhealth.9131] [Medline: 29549071]
22. Hawke LD, Sheikhan NY, MacCon K, Henderson J. Going virtual: youth attitudes toward and experiences of virtual mental health and substance use services during the COVID-19 pandemic. BMC Health Serv Res 2021 Apr 14;21(1):340 [FREE Full text] [doi: 10.1186/s12913-021-06321-7] [Medline: 33853602]
23. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
24. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
25. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd edition. Hoboken, NJ, USA: John Wiley & Sons; 2019.
26. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015 Jun 02;162(11):777-784 [FREE Full text] [doi: 10.7326/M14-2385] [Medline: 26030634]
27. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
28. Ryan RM, Deci EL. On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu Rev Psychol 2001;52:141-166. [doi: 10.1146/annurev.psych.52.1.141] [Medline: 11148302]
29. Keyes CL. The mental health continuum: from languishing to flourishing in life. J Health Soc Behav 2002 Jun;43(2):207-222. [Medline: 12096700]
30. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
31. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
32. Covidence: Better Systematic Review Management. Covidence. URL: https://www.covidence.org/ [accessed 2022-09-06]
33. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
34. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]

35. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
36. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-Analysis. Hoboken, NJ, USA: Wiley; 2009.
37. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Hoboken, NJ, USA: John Wiley & Sons; 2008.
38. Ponzo S, Morelli D, Kawadler J, Hemmings NR, Bird G, Plans D. Efficacy of the digital therapeutic mobile app BioBase to reduce stress and improve mental well-being among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 06;8(4):e17767 [FREE Full text] [doi: 10.2196/17767] [Medline: 31926063]
39. Levin ME, Hayes SC, Pistorello J, Seeley JR. Web-based self-help for preventing mental health problems in universities: comparing acceptance and commitment training to mental health education. J Clin Psychol 2016 Mar;72(3):207-225. [doi: 10.1002/jclp.22254] [Medline: 26784010]
40. Levin ME, Hicks ET, Krafft J. Pilot evaluation of the stop, breathe & think mindfulness app for student clients on a college counseling center waitlist.. J Am Coll Health 2020;70(1):165-173. [doi: 10.1080/07448481.2020.1728281]
41. Levin ME, Krafft J, Hicks ET, Pierce B, Twohig MP. A randomized dismantling trial of the open and engaged components of acceptance and commitment therapy in an online intervention for distressed college students. Behav Res Ther 2020 Mar;126:103557. [doi: 10.1016/j.brat.2020.103557] [Medline: 32014692]
42. Viskovich S, Pakenham KI. Randomized controlled trial of a Web-based Acceptance and Commitment Therapy (ACT) program to promote mental health in university students. J Clin Psychol 2020 Jun;76(6):929-951. [doi: 10.1002/jclp.22848] [Medline: 31468528]
43. Bendtsen M, Bendtsen P, Henriksson H, Henriksson P, Müssener U, Thomas K, et al. The mobile health multiple lifestyle behavior interventions across the Lifespan (MOBILE) research program: protocol for development, evaluation, and implementation. JMIR Res Protoc 2020 Apr 20;9(4):e14894 [FREE Full text] [doi: 10.2196/14894] [Medline: 32310147]
44. Viskovich S, Pakenham KI. Pilot evaluation of a Web-based acceptance and commitment therapy program to promote mental health skills in university students. J Clin Psychol 2018 Dec;74(12):2047-2069. [doi: 10.1002/jclp.22656] [Medline: 29962090]
45. Noone C, Hogan MJ. A randomised active-controlled trial to examine the effects of an online mindfulness intervention on executive control, critical thinking and key thinking dispositions in a university student sample. BMC Psychol 2018 Apr 05;6(1):13 [FREE Full text] [doi: 10.1186/s40359-018-0226-3] [Medline: 29622047]
46. Auyeung L, Mo PK. The efficacy and mechanism of online positive psychological intervention (PPI) on improving well-being among Chinese university students: a pilot study of the best possible self (BPS) intervention. J Happiness Stud 2019;20(8):2525-2550. [doi: 10.1007/s10902-018-0054-4]
47. Firestone J, Cardaciotto L, Levin ME, Goldbacher E, Vernig P, Gambrel LE. A Web-based self-guided program to promote valued-living in college students: a pilot study. J Context Behav Sci 2019 Apr;12:29-38. [doi: 10.1016/j.jcbs.2019.01.004]
48. Kvillemo P, Brandberg Y, Bränström R. Feasibility and outcomes of an Internet-based mindfulness training program: a pilot randomized controlled trial. JMIR Ment Health 2016 Jul 22;3(3):e33 [FREE Full text] [doi: 10.2196/mental.5457] [Medline: 27450466]
49. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017 Jan 01;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
50. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. Behav Res Ther 2016 Mar;78:30-42. [doi: 10.1016/j.brat.2016.01.001] [Medline: 26848517]
51. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208 [FREE Full text] [doi: 10.2196/17208] [Medline: 32196462]
52. Beerse ME, Van Lith T, Stanwood G. Therapeutic psychological and biological responses to mindfulness-based art therapy. Stress Health 2020 Oct;36(4):419-432. [doi: 10.1002/smi.2937] [Medline: 32073202]
53. Keyes CL, Wissing M, Potgieter JP, Temane M, Kruger A, van Rooy S. Evaluation of the mental health continuum-short form (MHC-SF) in setswana-speaking South Africans. Clin Psychol Psychother 2008;15(3):181-192. [doi: 10.1002/cpp.572] [Medline: 19115439]
54. Ryff CD. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. J Pers Soc Psychol 1989 Dec;57(6):1069-1081. [doi: 10.1037/0022-3514.57.6.1069]
55. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
56. Keyes CL, Shmotkin D, Ryff CD. Optimizing well-being: the empirical encounter of two traditions. J Pers Soc Psychol 2002 Jun;82(6):1007-1022. [Medline: 12051575]
57. 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 2009 May 28;97(2):143-156. [doi: 10.1007/s11205-009-9493-y]

58. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
59. Sheehan TJ, Fifield J, Reisine S, Tennen H. The measurement structure of the Center for Epidemiologic Studies Depression Scale. J Pers Assess 1995 Jun;64(3):507-521. [doi: 10.1207/s15327752jpa6403_9] [Medline: 7760258]
60. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]
61. Cohen S, Williamson GM. Perceived stress in a probability sample of the United States. In: Spacapam S, Oskamp S, editors. Social Psychology of Health. Newbury Park, CA, USA: Sage Publications; 1988:31-67.
62. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
63. Crawford JR, Garthwaite PH, Lawrie CJ, Henry JD, MacDonald MA, Sutherland J, et al. A convenient method of obtaining percentile norms and accompanying interval estimates for self-report mood scales (DASS, DASS-21, HADS, PANAS, and sAD). Br J Clin Psychol 2009 Jun;48(Pt 2):163-180. [doi: 10.1348/014466508X377757] [Medline: 19054433]
64. Bleich SN, Vercammen KA, Zatz LY, Frelier JM, Ebbeling CB, Peeters A. Interventions to prevent global childhood overweight and obesity: a systematic review. Lancet Diabetes Endocrinol 2018 Apr;6(4):332-346. [doi: 10.1016/S2213-8587(17)30358-3] [Medline: 29066096]
65. Gauntlett L, Amlôt R, Rubin GJ. How to inform the public about protective actions in a nuclear or radiological incident: a systematic review. Lancet Psychiatry 2019 Jan;6(1):72-80. [doi: 10.1016/S2215-0366(18)30173-1] [Medline: 30340985]
66. 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]
67. Duval S, Tweedie R. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. J Am Stat Assoc 2000 Mar;95(449):89-98. [doi: 10.1080/01621459.2000.10473905]
68. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change. Washington, DC, USA: American Psychological Association; 2009.
69. Harris R. ACT Made Simple: An Easy-To-Read Primer on Acceptance and Commitment Therapy. Oakland, CA, USA: New Harbinger Publications; 2019.
70. Hayes SC, Pistorello J, Levin ME. Acceptance and commitment therapy as a unified model of behavior change. Couns Psychol 2012 Sep 26;40(7):976-1002. [doi: 10.1177/0011000012460836]
71. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change. 2nd edition. New York, NY, USA: The Guilford Press; 2012.
72. Ochnik D, Rogowska AM, Kuśnierz C, Jakubiak M, Schütz A, Held MJ, et al. Mental health prevalence and predictors among university students in nine countries during the COVID-19 pandemic: a cross-national study. Sci Rep 2021 Sep 20;11(1):18644 [FREE Full text] [doi: 10.1038/s41598-021-97697-3] [Medline: 34545120]
73. Armour C, McGlinchey E, Butter S, McAloney-Kocaman K, McPherson K. The COVID-19 psychological wellbeing study: understanding the longitudinal psychosocial impact of the COVID-19 pandemic in the UK; a methodological overview paper. J Psychopathol Behav Assess 2021;43(1):174-190 [FREE Full text] [doi: 10.1007/s10862-020-09841-4] [Medline: 33169046]
74. Ali W. Online and remote learning in higher education institutes: a necessity in light of COVID-19 pandemic. Higher Educ Stud 2020 May 18;10(3):16. [doi: 10.5539/hes.v10n3p16]
75. Valentine L, McEnery C, Bell I, O'Sullivan S, Pryor I, Gleeson J, et al. Blended digital and face-to-face care for first-episode psychosis treatment in young people: qualitative study. JMIR Ment Health 2020 Jul 28;7(7):e18990 [FREE Full text] [doi: 10.2196/18990] [Medline: 32720904]
76. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
77. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):e20-e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
78. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]

## Abbreviations

**ACT:** acceptance and commitment therapy
**DASS:** Depression Anxiety Stress Scale
**HADS:** Hospital Anxiety and Depression Scale
**PICO:** population, intervention, control, and outcomes
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PWB:** psychological well-being

**RCT:** randomized controlled trial
**RoB:** risk of bias

---

*Edited by T Leung; submitted 18.05.22; peer-reviewed by A Hudon, S Goodday; comments to author 09.08.22; revised version received 16.08.22; accepted 18.08.22; published 28.09.22.*

*Please cite as:*
*Ferrari M, Allan S, Arnold C, Eleftheriadis D, Alvarez-Jimenez M, Gumley A, Gleeson JF*
*Digital Interventions for Psychological Well-being in University Students: Systematic Review and Meta-analysis*
*J Med Internet Res 2022;24(9):e39686*
*URL: https://www.jmir.org/2022/9/e39686*
*doi:10.2196/39686*
*PMID:36169988*

Review

# Enabling Early Obstructive Sleep Apnea Diagnosis With Machine Learning: Systematic Review

Daniela Ferreira-Santos[1,2], MSc; Pedro Amorim[1,2,3], MSc; Tiago Silva Martins[2], MSc; Matilde Monteiro-Soares[1,2,4], PhD; Pedro Pereira Rodrigues[1,2], PhD

[1]Department of Community Medicine, Information and Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal
[2]Center for Health Technology and Services Research, Porto, Portugal
[3]Sleep and Non-Invasive Ventilation Unit, São João University Hospital, Porto, Portugal
[4]Portuguese Red Cross Health School Lisbon, Lisbon, Portugal

**Corresponding Author:**
Daniela Ferreira-Santos, MSc
Department of Community Medicine, Information and Decision Sciences
Faculty of Medicine, University of Porto
Rua Dr Plácido da Costa, s/n
Porto, 4200-450
Portugal
Phone: 351 937710766
Email: danielasantos@med.up.pt

## Abstract

**Background:** American Academy of Sleep Medicine guidelines suggest that clinical prediction algorithms can be used to screen patients with obstructive sleep apnea (OSA) without replacing polysomnography, the gold standard.

**Objective:** We aimed to identify, gather, and analyze existing machine learning approaches that are being used for disease screening in adult patients with suspected OSA.

**Methods:** We searched the MEDLINE, Scopus, and ISI Web of Knowledge databases to evaluate the validity of different machine learning techniques, with polysomnography as the gold standard outcome measure and used the Prediction Model Risk of Bias Assessment Tool (Kleijnen Systematic Reviews Ltd) to assess risk of bias and applicability of each included study.

**Results:** Our search retrieved 5479 articles, of which 63 (1.15%) articles were included. We found 23 studies performing diagnostic model development alone, 26 with added internal validation, and 14 applying the clinical prediction algorithm to an independent sample (although not all reporting the most common discrimination metrics, sensitivity or specificity). Logistic regression was applied in 35 studies, linear regression in 16, support vector machine in 9, neural networks in 8, decision trees in 6, and Bayesian networks in 4. Random forest, discriminant analysis, classification and regression tree, and nomogram were each performed in 2 studies, whereas Pearson correlation, adaptive neuro-fuzzy inference system, artificial immune recognition system, genetic algorithm, supersparse linear integer models, and k-nearest neighbors algorithm were each performed in 1 study. The best area under the receiver operating curve was 0.98 (0.96-0.99) for age, waist circumference, Epworth Somnolence Scale score, and oxygen saturation as predictors in a logistic regression.

**Conclusions:** Although high values were obtained, they still lacked external validation results in large cohorts and a standard OSA criteria definition.

**Trial Registration:** PROSPERO CRD42021221339; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=221339

***(J Med Internet Res 2022;24(9):e39452)*** doi:10.2196/39452

**KEYWORDS**

machine learning; obstructive sleep apnea; systematic review; polysomnography

# Introduction

## Background

Obstructive sleep apnea (OSA) is a common sleep-related breathing disorder characterized by recurrent episodes of partial (hypopnea) or complete (apnea) upper airway obstruction, repeated throughout sleep. Its prevalence varies significantly according to how OSA is defined (methodology, criteria used such as apnea index, apnea-hypopnea index [AHI], or respiratory disturbance index and threshold definitions) and the population being studied [1]. The study by Benjafield et al [2] estimated that worldwide, 936 million adults aged 30 to 69 years have OSA. Despite this high prevalence, many cases remain undiagnosed and untreated, leading to a decrease in patients' quality of life and an increased risk of adverse events, with a high impact on morbidity and mortality [3]. Polysomnography (PSG) is the gold standard test for diagnosing OSA [1]. However, performing PSG is costly, time-consuming, and labor-intensive. Most sleep laboratories face long waiting lists of patients, as PSG is neither a routine clinical practice nor an absolute suitable screening tool [4]. Given these limitations, it would be useful to develop a clinical prediction model that could reliably identify the patients most likely to benefit from PSG, that is, exclude OSA diagnosis when the probability is low, establish a priori probability before considering PSG, and prioritize patients in need of PSG according to the probability of a positive result. This idea was backed up by the American Academy of Sleep Medicine (AASM) in its latest guidelines [1]. Clinical prediction models should be easy to use and easy to calculate. The model must be based on the gold standard and required to be validated, and when used for screening, its purpose depends on whether the path leads to a rule-out or rule-in approach. In the first case, we should have a high-sensitivity model, omitting the need to perform PSG in healthy patients. By contrast, if we chose a rule-in approach, a high-specificity model is needed to select patients with a high probability of having OSA, suitable for undergoing PSG.

## Objective

Given these shortcomings, this systematic review aimed to identify, gather, and analyze existing machine learning approaches that are being used for disease screening in adult patients with suspected OSA.

# Methods

This systematic review was carried out according to a protocol registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42021221339).

## Search Strategy and Selection Criteria

We searched all evidence available in the MEDLINE database (PubMed) and in Scopus and ISI Web of Knowledge published until June 2020 in English, French, Spanish, or Portuguese. Specific queries were used (with a refresh in October 2021), and a manual search was also performed by using the references of the included studies and pertinent reviews on the topic. In addition, contact with specialists in the field was made to check whether all pertinent information was retrieved. Articles were selected by 3 reviewers independently (blinded to each other's assessment) by applying the criteria to each title and abstract and then assessed fully. Divergent opinions were resolved through consensus. All processes were performed in Rayyan, a web application and mobile app for systematic reviews [5].

Studies including adult patients with suspected OSA (population) that assessed the accuracy of predictive models using known symptoms and signs of OSA (exposure and comparator) and had PSG as the gold standard (outcome) were eligible as per the selection criteria.

## Data Extraction

Once the articles were selected, data were extracted into a prespecified Excel spreadsheet and included (1) article information: title, author(s), publication date, country, and journal and (2) methods: study design, setting, study period, type of model, inclusion and exclusion criteria, participant selection, sample size, clinical factors analyzed, diagnostic test analyzed, and potential bias. For each type of model, specific data extraction was created and fulfilled, as demonstrated in the tables in further sections. We have ordered the identified studies by the obtained article results: first, the articles that only developed the algorithm; then the ones that internally validated the algorithm; and finally, the ones that externally validated the prediction algorithm. Within each subsection, we organized the published works by year of publication. Any missing information from the studies is reported in the Results section by "—" (not available), and the best obtained predictive model is marked in italic. Also, if the study applied different machine learning approaches, the clinical factors analyzed, and the discrimination measures are only described for the best obtained model.

## Risk of Bias

At 2 points in time, 1 reviewer assessed the risk of bias and applicability by applying the Prediction Model Risk of Bias Assessment Tool (PROBAST) to all the included studies. This is specific for studies developing, validating, or updating diagnostic prediction models. More details are available in the study by Moons et al [6]. An important aspect needs to be referred to, as this tool states that *"if a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains, consider downgrading to high risk of bias. Such a model can only be considered as low risk of bias if the development was based on a very large data set and included some form of internal validation."* This means that the included studies only performing model development will be marked as high risk of bias. For those with internal validation, the risk of bias will depend on the sample size based on the number of events per variable (≥20 ratio between events and variables in development studies and ≥100 participants with OSA for model validation studies). In addition, studies that randomly split a single data set into development and validation are considered as internal validation.

## Results

### Overview

We retrieved 6769 articles, 1290 being duplicates. From the 5479 articles, we kept 63 studies that fulfilled the inclusion criteria, as shown in Figure 1.

The gold-standard examination—PSG—was performed in all the articles assessed, with one also adding the diagnostic part of the split-night exam [7]. The highest found age was 96 years [8], with 54% (34/63) of studies presenting patients with ages of >18 years. To be certain to include all OSA clinical prediction algorithms, we kept the studies that only reported a mean age and SD, with this value being >42, and SD varying between 6 and 16 years. In addition, 10% (6/63) of studies reported an age group <18 years (>14 and >15 years in 2/6, 33% studies and >16 and >17 in 4/6, 66% others, respectively). Regarding the suspicion of OSA, this description was shown in 65% (41/63) of studies, whereas 32% (20/63) introduced OSA suspicion and any other sleep disorder. In addition, we have a study with healthy patients and patients with suspected OSA [9] and another that does not specifically state this; instead, the authors write that patients already diagnosed with OSA were excluded from the study. The frequency of occurrence of the various clinical factors analyzed in more than 1 study is shown in Table 1.

There were disagreements between the reviewers in both phases, with an overall concordance rate of 78% in the title and abstract screening and 95% in the integral version.

**Figure 1.** Flow diagram of the study selection process.

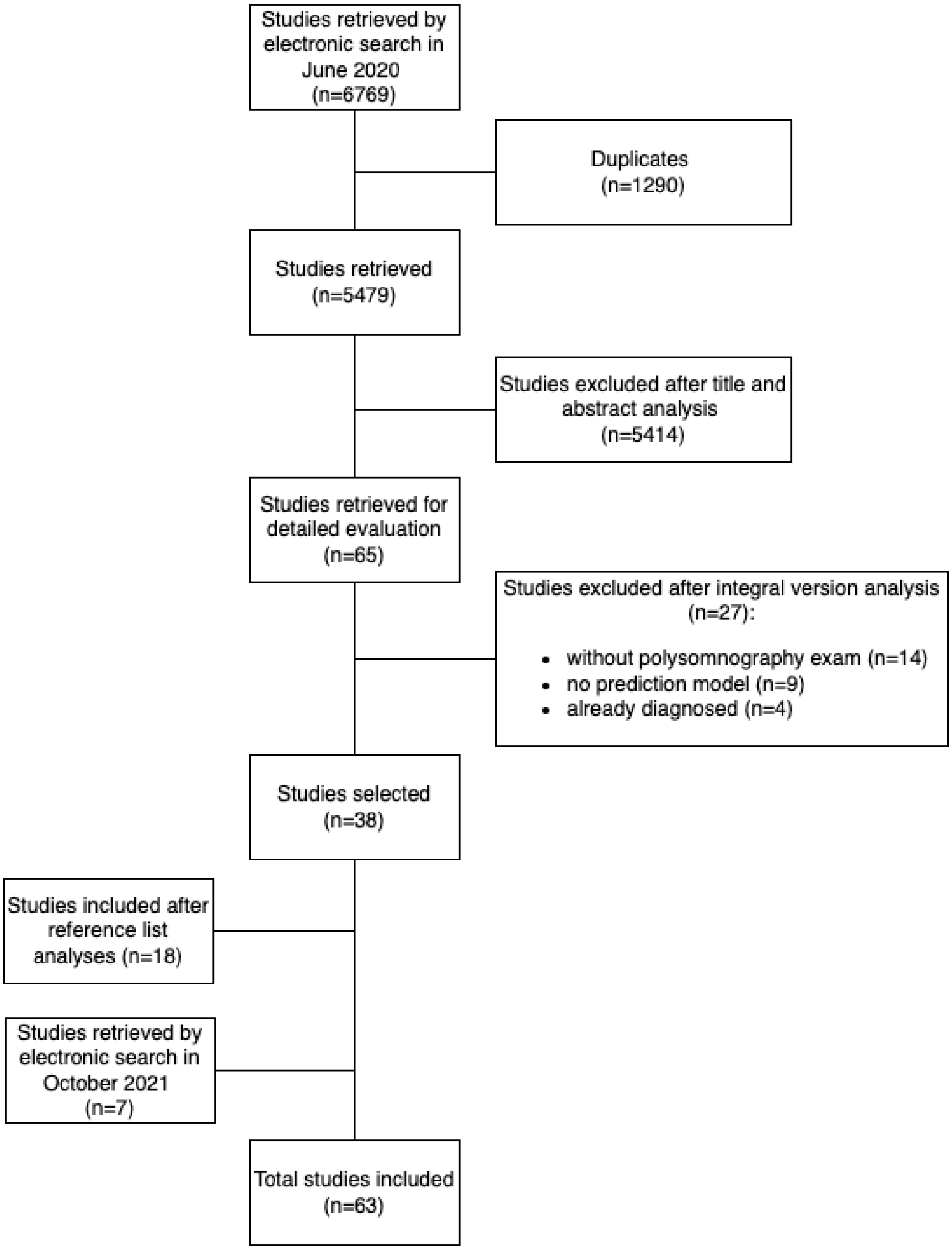


**Table 1.** The frequency of occurrence of the various clinical factors analyzed that appears more than once in all the included studies (n=63).

| Clinical factors analyzed | Frequency of occurrence, n (%) |
|---|---|
| BMI | 37 (59) |
| Age | 32 (51) |
| Sex | 29 (46) |
| Neck circumference | 25 (40) |
| Snoring | 14 (22) |
| Epworth Somnolence Scale | 10 (16) |
| Witnessed apneas | 8 (13) |
| Waist circumference | 8 (13) |
| Breathing cessation | 7 (11) |
| Daytime sleepiness | 7 (11) |
| Hypertension | 7 (11) |
| Gasping | 6 (10) |
| Oxygen saturation | 6 (10) |
| Oxygen desaturation | 6 (10) |
| Blood pressure | 5 (8) |
| Smoking | 5 (8) |
| Tonsil size grading | 5 (8) |
| Modified Mallampati score | 4 (6) |
| Alcohol consumption | 3 (5) |
| Awakenings | 3 (5) |
| Diabetes | 3 (5) |
| Height | 3 (5) |
| Nocturia | 3 (5) |
| Restless sleep | 3 (5) |
| Weight | 3 (5) |
| Craniofacial abnormalities | 2 (3) |
| Driving sleepy | 2 (3) |
| Face width | 2 (3) |
| Friedman tongue score | 2 (3) |
| Snorting | 2 (3) |

## Prediction Models Development

New prediction models were developed in 23 studies, as presented and described in Table 2. The most common approach was regression techniques, with logistic (6/23, 26%), linear (6/23, 26%), logistic and linear (6/23, 26%), and logistic regression compared with decision trees and support vector machines (3/23, 13%). In addition, 4% (1/23) of articles produced a Pearson correlation and another (1/23, 4%) produced a decision tree. The oldest model was developed in 1991 and included sex, age, BMI, and snoring whereas in 2020 the predictive variables included besides these were height, weight, waist size, hip size, neck circumference (NC), modified Friedman score, daytime sleepiness, and Epworth Somnolence Scale score. Only 13% (3/23) studies described the study design and period, with 22% (5/23) being retrospective. Regarding OSA definition by PSG, 4% (1/23) study did not report the cutoff, while 17% (4/23) reported an AHI>10 and 17% (4/23) more reported an AHI≥15. The largest sample size was 953, and the smallest was 96 patients with suspected OSA. An overall prevalence of OSA between 31% and 87% was stated, with 9% (2/23) of studies presenting incorrect percentage values [10,11]. Regarding discrimination measures, although no validation was performed, the best area under the receiver operating characteristic curve (AUC), sensitivity, and specificity were 99%, 100%, and 95%, respectively. It should also be noted that 4% (1/23) has no mention of the best prediction model (not marked in italic in Table 2).

**Table 2.** Studies' characteristics of prediction model development without internal or external validation with the best obtained model marked as italic in the respective model column.

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Viner et al [12], 1991 | Prospective; —[c] | Logistic regression | Sex, age, BMI, and snoring | AHI[d]>10 | 410 | 190 (46) | 77 (73-82) | 28 (—) | 95 (—) |
| Keenan et al [13], 1993 | — | Logistic regression | NC[e], age, WA[f], daytime sleepiness, driving sleepy, oxygen desaturation, and heart rate frequency | AHI>15 | 96 | 51 (53) | — | 20 (—) | 5 (—) |
| Hoffstein et al [14], 1993 | — | Linear regression | Subjective impression | AHI>10 | 594 | 275 (46) | — | 60 (—) | 63 (—) |
| Flemons et al [15] 1994 | —; February 1990 to September 1990 | Logistic and *linear* regression | NC, hypertension, snoring, and gasping or choking | AHI>10 | 175 | 82 (46) | — | — | — |
| Vaidya et al [16], 1996 | —; July 1993 to December 1994 | *Logistic* and linear regression | Age, BMI, sex, and total number of symptoms | RDI[g]>10 | 309 | 226 (73) | — | 96 (—) | 23 (—) |
| Deegan et al [11], 1996 | Prospective; — | Logistic and linear regression | Sex, age, snoring, WA, driving sleepy, alcohol consumption, BMI, number of dips ≥4%, lowest oxygen saturation, and NC | AHI≥15 | 250 | 135 (54) | — | — | — |
| Pradhan et al [17], 1996 | Prospective; August 1994 to February 1995 | Logistic regression | BMI, lowest oxygen saturation, and bodily pain score | RDI>10 | 150 | 85 (57) | — | 100 (—) | 31 (—) |
| Friedman et al [18], 1999 | Prospective; — | Linear regression | Modified Mallampati class, tonsil size grading, and BMI | RDI>20 | 172 | — | — | — | — |
| Dixon et al [19], 2003 | — | *Logistic* and linear regression | BMI, WA, glycosylated hemoglobin, fasting plasma insulin, sex, and age | AHI≥30 | 99 | 36 (36) | 91 (—) | 89 (—) | 81 (—) |
| Morris et al [10], 2008 | Prospective; — | Pearson correlation | BMI and snoring severity score | RDI≥15 | 211 | 175 (83) | — | 97 (—) | 40 (—) |
| Martinez-Rivera et al [20], 2008 | — | Logistic regression | Sex, waist-to-hip ratio, BMI, NC, and age | AHI>10 | 192 | 124 (65) | — | — | — |

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Herzog et al [21], 2009 | Retrospective; — | Logistic and *linear* regression | Tonsil size grading, uvula size, dorsal movement during simulated snoring, collapse at tongue level, BMI, and ESS[h] score | AHI>5 | 622 | — | — | Female: 98 (—) | Female: 22 (—) |
| Yeh et al [22], 2010 | Retrospective; April 2006 to December 2007 | Linear regression | BMI, NC, and ESS score | AHI≥15 | 101 | 83 (82) | — | 98 (—) | — |
| Hukins et al [23], 2010 | Retrospective; January 2005 to July 2007 | Linear regression | Mallampati class IV | AHI>30 | 953 | 297 (31) | — | 40 (36-45) | 67 (64-69) |
| Musman et al [24], 2011 | —; December 2006 to March 2007 | Logistic and *linear* regression | NC, WA, age, BMI, and allergic rhinitis | AHI>5 | 323 | 229 (71) | — | — | — |
| Sareli et al [25], 2011 | —; November 2005 to January 2007 | Logistic regression | Age, BMI, sex, and sleep apnea symptom score | AHI≥5 | 342 | 264 (77) | 80 (—) | — | — |
| Tseng et al [26], 2012 | — | Decision tree | Sex, age, preovernight systolic blood pressure, and postovernight systolic blood pressure | AHI≥15 | 540 | 394 (73) | — | — | — |
| Sahin et al [27], 2014 | Retrospective; — | Linear regression | BMI, WC[i], NC, oxygen saturation, and tonsil size grading | AHI>5 and symptoms | 390 | — | — | — | — |
| Ting et al [28], 2014 | Prospective; — | Logistic regression and *decision trees* | Sex, age, and blood pressure | AHI≥15 | 540 | 394 (73) | 99 (—) | 98 (—) | 93 (—) |
| Sutherland et al [29], 2016 | —; 2011 to 2012 | *Logistic* regression and classification and regression tree | Face width and cervicomental angle | AHI≥10 | 200 | 146 (73) | 76 (68-83) | 89 (—) | 28 (—) |
| Lin et al [4], 2019 | Retrospective; — | Linear regression | Sex, updated Friedman tongue position, tonsil size grading, and BMI | AHI≥5 | 325 | 283 (87) | 80 (74-87) | 84 (—) | 58 (—) |
| Del Brutto et al [30], 2020 | — | Logistic regression | Neck grasp | AHI≥5 | 167 | 114 (68) | 62 (54-69) | 83 (75-89) | 40 (27-54) |

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Haberfeld et al [8], 2020 | — | Logistic regression and *support vector machine* | Height, weight, WC, hip size, BMI, age, neck size, modified Friedman score, snoring, sex, daytime sleepiness, and ESS score | — | 620 | 357 (58) | Male: 61 (—) | Male: 86 (—) | Male: 70 (—) |

[a] OSA: obstructive sleep apnea.

[b] AUC: area under receiver operating characteristic curve.

[c] Not available.

[d] AHI: apnea-hypopnea index.

[e] NC: neck circumference.

[f] WA: witnessed apnea.

[g] RDI: respiratory disturbance index.

[h] ESS: Epworth somnolence scale.

[i] WC: waist circumference.

As stated in the Methods section, given that all these models only performed development with in-sample validation metrics, they were all considered at high risk of bias in the Analysis domain (Table 3). Concerning the Outcome domain, most studies were marked as high risk, as most of them did not have a prespecified or standard outcome definition. In addition, although some were marked as high risk and one as unclear, most included studies were at low risk of bias regarding the Predictors domain, showing that most of the studies did not include predictors after performing PSG. Most studies (15/23, 65%) were identified as unclear for the Participants domain, as almost all studies did not state study design or exclusion criteria. Assessing the applicability aspect of PROBAST, all studies (23/23, 100%) were at low risk of bias for the Participants domain (all studies included patients with suspected OSA), but several were at high risk of applicability for the Outcome domain (OSA definition is not in concordance with current OSA guidelines).

**Table 3.** Prediction Model Risk of Bias Assessment Tool (PROBAST) for prediction model development without internal or external validation.

| Study | Risk of bias | | | | Applicability | | | Overall | |
|---|---|---|---|---|---|---|---|---|---|
| | Participants | Predictors | Outcome | Analysis | Participants | Predictors | Outcome | Risk of bias | Applicability |
| Viner et al [12], 1991 | ☒ [a] | ☒ [b] | ☒ [c] | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Keenan et al [13], 1993 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Hoffstein et al [14], 1993 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Flemons et al [15], 1994 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Vaidya et al [16], 1996 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Deegan et al [11], 1996 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Pradhan et al [17], 1996 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Friedman et al [18], 1999 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Dixon et al [19], 2003 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Morris et al [10], 2008 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Martinez-Rivera et al [20], 2008 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Herzog et al [21], 2009 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Yeh et al [22], 2010 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Hukins [23], 2010 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Musman et al [24], 2011 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Sareli et al [25], 2011 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Tseng et al [26], 2012 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Sahin et al [27], 2014 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Ting et al [28], 2014 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Sutherland et al [29], 2016 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Lin et al [4], 2019 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Del Brutto et al [30], 2020 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Haberfeld et al [8], 2020 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |

[a] Indicates an unclear risk of bias or concerns regarding applicability.

[b] Indicates a low risk of bias or concerns regarding applicability.

[c] Indicates a high risk of bias or concerns regarding applicability.

### Development of Prediction Models With Internal Validation

For purposes of internal validation, we considered studies that performed cross-validation (11/26, 42%), used bootstrapping techniques (4/26, 15%), or used split-data (14/26, 54%) as previously mentioned in the Methods section. The smallest sample size was 83 participants and the highest was 6399, with both presenting validation results for cross-validation. Regarding OSA prevalence, a study had no mention, and another demonstrated an incorrect value [31], whereas others had the lowest value at 30% and the highest at 90%. Different machine learning approaches were used, with the most common being support vector machines (4/26, 15%), followed by logistic regression (3/26, 12%). Moreover, 38% (10/26) of studies described the study type and period, with retrospective design being the most common.

In addition, Table 4 shows different OSA definitions, with 8% (2/26) of studies not reporting cutoff values and the most common definition being AHI≥5 (8/26, 31%), followed by AHI≥15 (5/26, 19%). It should be noted that although the studies indicated that some types of internal validation were performed, some did not present results (10/26, 38%).

Regarding discrimination measures for internal validation, the best AUC, sensitivity, and specificity were 97%, 99%, and 97%, respectively. The model with the best AUC included predictive variables collected from PSG, such as the arousal index, and was also the model with the best specificity. The best sensitivity value was obtained for the neural network model with 19 predictive variables included. A total of 4 studies reported a clinical cutoff, which allows potential clinical threshold importance, with 50% reported in 2 studies and 32% in the other two.

In contrast to Table 3, Table 5 demonstrated that although internal validation was performed, only 8% (2/26) of studies had a low risk of bias in the Analysis domain, the reason being not presenting the relevant calibration or discrimination measures, such as AUC, and using only *P* values to select predictors. Furthermore, in the Participants domain applicability, 8% (2/26) of studies were marked as having a high risk of applicability, as they did not select only patients with suspected OSA.

**Table 4.** Studies' characteristics of prediction model development with internal validation. If the study applied different machine learning approaches, the clinical factors analyzed and the discrimination measures are only described for the best obtained model, marked as italic in the respective model column.

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Kapuniai et al [9], 1988 | —[c] | Discriminant analysis | Breathing cessation, adenoidectomy, BMI, and gasping | AI[d]>5 | $D_1$[e]=43; $D_2$=53 | 13 (30) | — | 61 (—) | 67 (—) |
| Kirby et al [32], 1999 | Retrospective; — | Neural network | Age, sex, frequent awakening, experienced choking, WA[f], observed choking, daytime sleepiness, ESS[g], hypertension, alcohol consumption, smoking, height, weight, BMI, blood pressure, tonsillar enlargement, soft-palate enlargement, crowding of the oral pharynx, and sum of the clinical scores for the binary categorical values | AHI[h]≥10 | $D_1$=255; $D_2$=150 | 281 (69) | 94 (—) | 99 (97-100) | 80 (70-90) |
| Lam et al [33], 2005 | Prospective; January 1999 to December 1999 | Discriminant analysis | Mallampati score, thyromental angle, NC[i], BMI, age, and thyromental distance | AHI≥5 | $D_1$=120; $D_2$=119[j] | 201 (84) | 71 (—)[k] | — | — |
| Julià-Serdà et al [34], 2006 | — | Logistic regression | NC, sex, desaturation, ESS score, and distance between the gonion and the gnathion | AHI≥10 | $D_1$=150; $D_2$=57 | 115 (56) | 97 (95-99)[k] | 94 (—) | 83 (—) |
| Polat et al [35], 2008 | Prospective; — | *Decision tree*, neural network, 21 adaptive neuro-fuzzy inference system, and artificial immune recognition system | Arousals index, AHI, minimum oxygen saturation value in stage REM[l], and percentage of sleep time in stage of oxygen saturations intervals bigger than 89% | AHI>5 | $D_1$=41; $D_2$=42[j] | 58 (70) | 97 (—) | 92 (—) | 97 (—) |
| Chen et al [31], 2008 | —; January 2004 to December 2005 | Support vector machine | Oxygen desaturation index | AHI≥5 | 566[j] | 491 (87) | — | 43 (—) | 94 (—) |

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Lee et al [36], 2009 | Prospective; — | *Logistic regression* and classification and regression tree | Face width, eye width, mandibular length, WA, and modified Mallampati class | AHI≥10 | 180[j] | 114 (63) | 87 (—)[k] | 85 (—)[k] | 70 (—)[k] |
| Rofail et al [37], 2010 | —; July 2006 to November 2007 | Logistic regression | Index 1 (snoring, breathing cessation, snorting, gasping), and nasal flow RDI[m] | AHI≥5 | $D_1$=96; $D_2$=97 | 139 (72) | 89 (81-97) | 85 (—) | 92 (—) |
| Chen et al [38], 2011 | Retrospective; — | Logistic regression | Desaturation 3% | RDI≥30 | $D_j$=355; $D_2$=100[j] | 307 (86) | 95 (—)[k] | 90 (—) | 90 (—) |
| Bucca et al [39], 2011 | Prospective; January 2004 to December 2005 | Linear regression | Age, NC, BMI, FEF50/FIF50[n], $COH_B$%[o], smoking, $F_{eNO}$[p], and interaction smoking and $F_{eNO}$ | AHI≥30 | 201[q] | 120 (60) | — | — | — |
| Bouloukaki et al [40], 2011 | Prospective; October 2000 to December 2006 | Linear regression | NC, sleepiness severity, BMI, and sex | AHI≥15 | $D_1$=538; $D_2$=2152 | 2130 (79) | 78 (61-80)[k] | 70 (—)[k] | 73 (—)[k] |
| Sun et al [41], 2011 | —; February 2009 to June 2009 | Logistic regression and *genetic algorithm* | Demographic data, ESS, systemic diseases, snoring, and comorbidities | AHI≥15 | $D_1$=67; $D_2$=43 | 53 (48) | — | 82 (—) | 95 (—) |
| Laporta et al [42], 2012 | Prospective; October 2010 to September 2011 | Neural network | Age, weight, sex, height, NC, hypertension, daytime sleepiness, difficulty falling asleep, snoring, breathing cessation, restless sleep, and gasping | AHI≥5 | 91[q] | 68 (75) | 93 (85-97)[k] | 99 (92-100)[k] | 87 (66-97)[k] |
| Hang et al [43], 2013 | Retrospective; January 2005 to December 2006 | Support vector machine | Oxygen desaturation index, ESS, or BMI | AHI≥15 | $D_1$=188; $D_2$=188; $D_3$=189 | — | — | 88 (85-90)[k] | 90 (87-94)[k] |
| Hang et al [44], 2015 | —; January 2004 to December 2005 | Support vector machine | Oxygen desaturation index | AHI>30 | 1156[j] | 285 (46) | $D_1$: 96 (—)[k]; $D_2$: 95 (—)[k] | $D_1$: 87 (—); $D_2$: 91 (—)[k] | $D_1$: 93 (—); $D_2$: 90 (—)[k] |
| Ustun et al [7], 2016 | —; January 2009 to June 2013 | Logistic regression, *super-sparse linear integer models*, decision tree, and support vector machines | Age, sex, BMI, diabetes, hypertension, and smoking | AHI>5 | 1922[j] | 1478 (77) | 79 (—) | 64 (—) | 23 (—) |

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Bozkurt et al [45], 2017 | Retrospective; January 2014 to August 2015 | Logistic regression, *Bayesian network*, decision tree, random forest, and neural network | Sex, age, BMI, NC, and smoking | AHI≥5 | 338[j] | 304 (90) | 73 (—) | 86 (—) | 85 (—) |
| Ferreira-Santos [46], 2017 | Retrospective; January 2015 to May 2015 | Bayesian network | Sex, NC, CFA[r], WA, nocturia, alcohol consumption, ESS, concentration decrease, atrial fibrillation, stroke, myocardial infarction, driver, and daytime sleepiness | AHI≥5 | 194[j] | 128 (66) | 76 (73-78) | 81 (79-83) | 48 (44-51) |
| Liu et al [47], 2017 | —; October 2005 to April 2014 and October 2013 to September 2014 | Support vector machine | WC[s], NC, BMI, and age | AHI≥15 | 6399[j] | 3866 (60) | Female: 90 (87-94) | Female: 83 (75-91) | Female: 86 (82-90) |
| Manoochehri et al [48], 2018 | —; 2012 to 2016 | Logistic regression and *decision tree* | WC, snoring, sex, sleep apnea, ESS score, and NC | — | $D_1$=239; $D_2$=99 | 208 (62) | — | 67 (—) | 81 (—) |
| Manoochehri et al [49], 2018 | —; 2012 to 2015 | Logistic regression and *support vector machine* | Age, sex, BMI, NC, WC, tea consumption, smoking, hypertension, chronic headache, heart disease, respiratory disease, neurological disease, and diabetes | — | $D_1$=176; $D_2$=74 | 154 (62) | — | 71 (—)[k] | 85 (—)[k] |
| Xu et al [50], 2019 | —; 2007 to 2016 | Nomogram | Age, sex, glucose, apolipoprotein B, insulin, BMI, NC, and WC | AHI>5 | 4162[q] | 3387 (81) | 84 (83-86) | 77 (76-79)[k] | 76 (72-80)[k] |
| Ferreira-Santos et al [51], 2019 | Retrospective; January 2015 to May 2015 | Bayesian network | Sex, WA, age, nocturia, CFA, and NC | AHI≥5 | 194[j] | 128 (66) | 64 (61-66) | 90 (88-92) | 24 (20-27) |
| Keshavarz et al [52], 2020 | Retrospective; February 2013 to December 2017 | Logistic regression, Bayesian network, *neural network*, k-nearest neighbors, support vector machine, and random forest | Snoring, nocturia, awakening owing to the sound of snoring, snoring, back pain, restless sleep, BMI, and WA | AHI>15 | 231[j] | 152 (66) | 75 (—) | 86 (—) | 53 (—) |
| Chen et al [53], 2021 | Retrospective; September 2015 to January 2020 | Nomogram | Age, sex, snoring, type 2 diabetes mellitus, NC, and BMI | AHI≥5 | $D_1$=338; $D_2$=144[q] | 342 (71) | 83 (76-90) | 69 (63-75)[k] | 87 (79-93)[k] |

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Hsu et al [54], 2021 | —; December 2011 to August 2018 | *Logistic regression*, support vector machine, and neural network | Sex, age, and BMI | AHI≥15 | $D_1$=2446;<br>$D_2$=1049 | 2539 (73) | 82 (—) | 73 (—)[k] | 77 (—)[k] |

[a] OSA: obstructive sleep apnea.

[b] AUC: area under receiver operating characteristic curve.

[c] Not available.

[d] AI: apnea index.

[e] $D_1$, $D_2$, and $D_3$: data set.

[f] WA: witnessed apnea.

[g] ESS: Epworth somnolence scale.

[h] AHI: apnea-hypopnea index.

[i] NC: neck circumference.

[j] cross-validation.

[k] Internal derivation results.

[l] REM: rapid eye movement.

[m] RDI: respiratory disturbance index.

[n] FEF50/FIF50: forced midexpiratory/midinspiratory airflow ratio.

[o] COHB%: carboxyhemoglobin percent saturation.

[p] $Fe_{NO}$: exhaled nitric oxide.

[q] Bootstrapping.

[r] CFA: craniofacial and upper airway.

[s] WC: waist circumference.

**Table 5.** Prediction Model Risk of Bias Assessment Tool (PROBAST) for prediction model development with internal validation.

| Study | Risk of bias | | | | Applicability | | | Overall | |
|---|---|---|---|---|---|---|---|---|---|
| | Participants | Predictors | Outcome | Analysis | Participants | Predictors | Outcome | Risk of bias | Applicability |
| Kapuniai et al [9], 1988 | ☒ [a] | ☒ [b] | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Kirby et al [32], 1999 | ☒ | ☒ [c] | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Lam et al [33], 2005 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Julià-Serdà et al [34], 2006 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Polat et al [35], 2008 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Chen et al [31], 2008 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Lee et al [36], 2009 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Rofail et al [37], 2010 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Chen et al [38], 2010 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Bucca et al [39], 2010 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Bouloukaki et al [40], 2011 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Sun et al [41], 2011 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Laporta et al [42], 2012 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Hang et al [43], 2015 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Hang et al [44], 2015 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Ustun et al [7], 2016 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Bozkurt et al [45], 2017 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Ferreira-Santos et al [46], 2017 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Liu et al [47], 2017 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Manoochehri et al [48], 2018 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Manoochehri et al [49], 2018 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Xu et al [50], 2019 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Ferreira-Santos et al [51], 2019 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Keshavarz et al [52], 2020 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Chen et al [53], 2021 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Hsu et al [54], 2021 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |

[a] Indicates an unclear risk of bias or concerns regarding applicability.

[b] Indicates a high risk of bias or concerns regarding applicability.

[c] Indicates a low risk of bias or concerns regarding applicability.

### Development of Prediction Models With External Validation

A total of 12 studies performed external validation, as described in Table 6, with 9 (75%) of them choosing logistic regression for the machine learning approach. The other 25% (3/12) elected linear regression, neural networks, or both. Regarding the study design, 3 (25%) studies elected a prospective design for testing and validation and 8% (1/12) of studies for only validation. Similar to the studies that only performed internal validation, the lowest OSA prevalence was 30%, and the highest was 93%, with a sample size varying between 169 and 3432 participants with suspected OSA. The best discriminatory model was logistic regression; it included age, waist circumference, ESS, and minimum oxygen saturation, with an AUC of 0.98 (0.96-0.99), for an OSA definition of AHI≥5. The higher reached sensitivity (100%) was also for a logistic regression but for a cutoff of AHI≥15, including specific respiratory conductance and daytime arterial oxygen saturation. The study also presented a clinical cutoff of 50%. Concerning specificity, the value of 94% was the highest for an AI>10, with self-reporting apneas, NC index, age, and tendency to fall asleep unintentionally as predictive variables.

As shown in Table 7, which aggregates information from the test and validation data sets, most studies were marked as unclear risk of bias in the Participants domain, as the studies referred to the study design for the test population but not for the validation data set. In addition, only 17% (2/12) of studies had a high risk of bias for the Predictors domain, given that the predictors could take time to be assessed or collected. Regarding the Analysis domain, half (6/12, 50%) of the studies were marked as having a low risk of bias, with 33% (4/12) of studies not presenting adequate performance metrics. The applicability in the Predictors domain is unclear in 8% (1/12) of studies, as we cannot assess whether the predictors are available in primary health care.

**Table 6.** Studies' characteristics of prediction model development with external validation. If the study applied different machine learning approaches, the clinical factors analyzed and the discrimination measures are only described for the best obtained model, marked as italic in the respective model column.

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Crocker et al [55], 1990 | —[c]; October 1986 to May 1988 | Logistic regression | Age, breathing cessation, BMI, and hypertension | AHI[d]>15 | T[e]=100;<br>V[f]=105 | 62 (30) | — | 92 (—) | 51 (—) |
| Pillar et al [56], 1992 | — | Logistic regression | WA[g], NC[h] index, age, daytime and sleepiness | AI[i]>10 and symptoms | *t*=86;<br>$V_1$=50;<br>$V_2$=105 | — | — | $V_1$=88 (—);<br>$V_2$=32 (—) | $V_1$=25 (—);<br>$V_2$=94 (—) |
| Maislin et al [57], 1995 | — | Logistic regression | BMI, age, sex, index 1 (snoring, breathing cessation, snorting, and gasping), and BMI index 1 | RDI[j]≥10 | *t*=658;<br>V=193 | 760 (89) | 79 (—)[k] | — | — |
| Kushida et al [58], 1997 | Prospective; 6 months (V) | Linear regression | Palatal height, maxillary intermolar distance, mandibular intermolar distance, overjet, BMI, and NC | RDI≥5 | *t*=30;<br>V=300[l, m] | 254 (85) | 100 (—)[k] | 98 (95-99)[k] | 100 (92-100)[k] |
| El-Solh et al [59], 1999 | Retrospective (T) and prospective (V); November 1995 to December 1996 | *Neural network* and linear regression | Breathing cessation, restless sleep, decreased libido, disturbs bed partner, daytime sleepiness, restless legs, BMI, NC, age, gasping, snoring, and blood pressure | AHI>10 | *t*=189[l];<br>V=80 | 182 (68) | 96 (93-96) | 95 (90-98)[k] | 65 (50-78)[k] |
| Zerah-Lancner et al [60], 2000 | Retrospective (T) and prospective (V); — | Logistic regression | Specific respiratory conductance and daytime arterial oxygen saturation | AHI≥15 | *t*=168;<br>V=101 | 147 (55) | — | 100 (—) | 84 (—) |
| Rodsutti et al [61], 2004 | Prospective; February 2001 to April 2003 | Logistic regression | Age, sex, BMI, and breathing cessation | AHI≥5 | *t*=837;<br>V=243 | 569 (53) | 79 (—) | — | — |
| Khoo et al [62], 2011 | —; December 2005 to December 2007 and March 2008 to June 2008 | Logistic regression | Sex, age, NC, and frequent awakening with unrefreshing sleep | AHI≥20 | *t*=117;<br>V=52 | 77 (66) | 69 (—)[k] | 78 (—) | 45 (—) |
| Zou et al [63], 2013 | Retrospective; January 2007 to July 2011 | Logistic regression | Age, WC[n], ESS[o], and minimum oxygen saturation | AHI≥5 | *t*=2052;<br>V=784 | 2451 (87) | 98 (96-99) | 94 (92-96) | 86 (79-91) |
| Karamanli et al [64], 2016 | Retrospective; — | Neural network | Sex, age, BMI, and snoring | AHI≥10 | *t*=201;<br>V=15 | 140 (70) | — | — | — |

| Study | Study design; study period | Machine learning approach | Clinical factors analyzed | OSA[a] definition | Sample size, n | OSA prevalence, n (%) | AUC[b], % (95% CI) | Sensitivity, % (95% CI) | Specificity, % (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| Tawaranurak et al [65], 2020 | Prospective; June 2018 to June 2020 | Logistic regression | Sex, choking or apnea, blood pressure, NC, WC, and BMI | AHI≥15 | t=892; V=374 | 826 (93) | 75 (—)[k] | 93 (89-96) | 26 (18-35) |
| Park et al [66], 2021 | —; January 2011 to December 2018 | Logistic regression | Age, sex, BMI, hypertension, Berlin questionnaire score, and tonsil grade | AHI≥5 | t=2516; V=916 | — | 84 (—) | 78 (—) | 76 (—) |

[a] OSA: obstructive sleep apnea.

[b] AUC: area under receiver operating characteristic curve.

[c] Not available.

[d] AHI: apnea-hypopnea index.

[e] T: test data set.

[f] V: validation data set.

[g] WA: witnessed apnea.

[h] NC: neck circumference.

[i] AI: apnea index.

[j] RDI: respiratory disturbance index.

[k] Internal derivation results.

[l] Cross-validation.

[m] Bootstrapping.

[n] WC: waist circumference.

[o] ESS: Epworth Somnolence Scale.

**Table 7.** Prediction Model Risk of Bias Assessment Tool (PROBAST) for prediction model development with external validation.

| Study | Risk of bias | | | | Applicability | | | Overall | |
|---|---|---|---|---|---|---|---|---|---|
| | Participants | Predictors | Outcome | Analysis | Participants | Predictors | Outcome | Risk of bias | Applicability |
| Crocker et al [55], 1990 | ☒ [a] | ☒ [b] | ☒ [c] | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Pillar et al [56], 1994 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Maislin et al [57], 1995 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Kushida et al [58], 1997 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| El-Solh et al [59], 1999 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Zerah-Lancner et al [60] 2000 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Rodsutti et al [61], 2003 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Khoo et al [62], 2011 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Zou et al [63], 2013 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Karamanli et al [64], 2016 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Tawaranurak et al [65], 2021 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |
| Park et al [66], 2021 | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ | ☒ |

[a] Indicates an unclear risk of bias or concerns regarding applicability.

[b] Indicates a low risk of bias or concerns regarding applicability.

[c] Indicates a high risk of bias or concerns regarding applicability.

### Prediction Models With External Validation

A total of 2 studies [67,68], one in 2000 and another in 2006, performed the external validation of 5 prediction models. The first was a prospective study that evaluated 4 clinical prediction models [12,15,55,57] for predicting the presence of OSA (AHI≥10). They included 370 patients with suspected OSA who underwent PSG between July 1996 and October 1997. The achieved prevalence of OSA was 67%, and the results are shown in Figure 1 and Table 4 of the original article [67]. The highest AUC, sensitivity, and specificity reached were 74%, 96%, and 54%, respectively. The second study used 80 patients with suspected OSA to evaluate the model described in the study by Kushida et al [58]. The objective was to evaluate the clinical applicability and define a clinical cutoff to differentiate OSA severities. Although the authors stated that the clinical applicability exists, they could not define a threshold for clinical use, and they did not present any discrimination measures.

The study of Flemons et al [15], in addition to producing a new prediction model, also applied the 2 equations from studies by Crocker et al [55] and Viner et al [12] to the obtained data set. Although no actual values were presented, the authors stated that the AUCs were very similar.

Furthermore, the study by Flemons et al [15] was externally validated by Khoo et al [62], with 52 patients with suspected OSA, reaching an AUC of 69%. If a clinical threshold of 60% is defined, the model in this independent sample reached 78% sensitivity and 45% specificity.

## Discussion

### Principal Findings

The AASM guidelines [1] explicitly state that *“clinical prediction algorithms may be used in sleep clinic patients with suspected OSA but are not necessary to substitute the need for PSG,”* whereas *“in non-sleep clinic settings, these tools may be more helpful to identify patients who are at increased risk for OSA.”* The evaluation of these tools in a nonsleep clinic setting was not tackled by AASM experts, as it was beyond the guideline scope. Therefore, our work aimed to answer this question by complementing step 1 in the clinical algorithm developed for clinical suspicion of OSA using clinical prediction algorithms in a nonsleep setting. With this, we hope to estimate the probability that OSA is present in a population with suspected OSA that is not yet diagnosed by aggregating information from multivariable prediction models, stating the ones that are best at rule out and rule in.

As such, the studies that only developed a model are the ones that need to gather evidence on whether the model would be helpful to put into clinical practice (high overfitting). To do so, it is needed to validate the model in a new population data set. One way to do this is by splitting the data set or performing a validity assessment using different techniques, such as cross-validation or bootstrapping, or even better, by applying the algorithm to an independent sample.

Of the 63 included studies, only 14 (22%) performed both development and external validation or only external validation of the algorithm. Most selected studies only developed 36% (23/63) or developed and internally validated 41% (26/63) of prediction models.

The study by Zerah-Lancner et al [60] emerged as the best at rule-out OSA, described a sensitivity value of 100% for an OSA definition of AHI≥15. The predictive variables included were respiratory conductance and oxygen saturation, chosen from an external population of 101 participants. The best at rule-in OSA was the study by Pillar et al [56]; for a validation population of 155 participants, it demonstrated a specificity of 94% for an AI≥10 symptoms, with witnessed apneas, NC, age, and falling asleep easily as predictive variables. Both studies used logistic regression as the machine learning approach. The study by Kushida et al [58] reached maximum specificity, but the authors did not describe whether the obtained results were for testing or external validation, in a 300-participant validation data set. These 2 best models [56,60] were developed and validated in 2000 and 1992, respectively, and presented a high risk of bias and applicability, with none of the studies providing the discriminatory power of the model or metric CIs.

The most recent study by Park et al [66], performed in 2021 with a validation data set of 916 participants (largest sample), only reached values of 78% and 76% for sensitivity and specificity, respectively, when compared with the 2 previous best models. This was also a logistic regression, electing BMI, age, sex, Berlin questionnaire score, and tonsil grade as the clinical factors for an OSA definition of AHI≥5. Although this study continued to lack the reporting of study design or prevalence of OSA, it presented a low risk of bias and applicability. But it only included Asian patients, so it cannot be race generalized, as the authors mention.

## Strengths and Limitations

It is important to consider some of the limitations and strengths of our methods and those of the included clinical studies. Although we cannot be sure that we retrieved all published literature, we are confident that our methodology is adequate. Risk was minimized by performing the search in 3 search engines (1 related to health sciences and 2 others with broader spectrums) and in 2 periods.

The PROBAST demonstrated that we face a high risk of bias and applicability, even when only assessing external validation results. Almost all the studies do not report the study design, which can raise problems in generating absolute probabilities or even in terms of inappropriately including or excluding participants. In addition, the definition and measurement of predictors and their association with the outcome were high in the 2 studies, as some of the predictors were not available when the model was intended to be used. Although all outcome definitions were based on PSG, some did not report how the measure was calculated or selected different cutoff values than the ones described in the guidelines. While all studies used appropriate statistical analysis, some lacked a reasonable number of participants with the outcome, in the test or validation data sets. Information regarding exclusion criteria or handling of missing data was not described, and most studies selected predictors based on univariable analysis. Besides all participants who underwent the gold standard exam, some did not have suspected OSA as the only inclusion criterion.

Different approaches have been followed since 1988 with the aim of predicting whether OSA is present in an individual, contributing to unlocking the bottleneck of in-hospital screening or diagnosis. However, assessing the bias or applicability of these approaches is not an easy task, with only 3 studies presenting an overall low risk of bias and applicability [63,65,66]. Furthermore, common missing points need to be pointed out are (1) most studies did not report the study design or period; (2) OSA definition differed within time, guidelines, and studies; (3) OSA prevalence varied from 30% to 93%, with some studies not describing the proportion; (4) needed measures to assess diagnostic value such as sensitivity, specificity, and AUC are not reported, and when reported, did not present CIs; and (5) some studies only create the predictive model and others add the validation task, but external validation is still lacking in all the studies.

Regarding the chosen machine learning approaches, the most common was logistic regression (35/63, 56%), followed by linear regression (16/63, 25%), support vector machine (9/63, 14%), neural networks (8/63, 13%), decision trees (8/63, 13%), Bayesian networks (4/63, 6%), random forest (2/63, 3%), discriminant analysis (2/63, 3%), classification and regression tree (2/63, 3%), nomogram (2/63, 3%), Pearson correlation (1/63, 2%), adaptive neuro-fuzzy inference system (1/63, 2%), artificial immune recognition system (1/63, 2%), genetic algorithm (1/63, 2%), supersparse linear integer models (1/63, 2%), and the k-nearest neighbors algorithm (1/63, 2%).

## Conclusions

In summary, this review provides an extensive, comprehensive, and up-to-date synthesis of diagnostic models in OSA. It is possible to predict OSA by only taking into consideration simple and available predictors such as BMI, age, sex, or NC as well as by reaching high levels of sensitivity or specificity, depending on whether we want to elect a rule-out or rule-in approach.

## Acknowledgments

DFS acknowledges Fundação para a Ciência e Tecnologia under PhD grants (PD/BD/13553/2018 and COVID/BD/152608/2022) for funding. This paper was supported by National Funds through Fundação para a Ciência e a Tecnologia, I.P., within the Center for Health Technology and Services Research, Research and Development Unit (reference UIDP/4255/2020).

## Conflicts of Interest

None declared.

## References

1. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American academy of sleep medicine clinical practice guideline. J Clin Sleep Med 2017 Mar 15;13(3):479-504 [FREE Full text] [doi: 10.5664/jcsm.6506] [Medline: 28162150]
2. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MS, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respiratory Med 2019 Aug;7(8):687-698. [doi: 10.1016/s2213-2600(19)30198-5]
3. Kendzerska T, Mollayeva T, Gershon AS, Leung RS, Hawker G, Tomlinson G. Untreated obstructive sleep apnea and the risk for serious long-term adverse outcomes: a systematic review. Sleep Med Rev 2014 Feb;18(1):49-59. [doi: 10.1016/j.smrv.2013.01.003] [Medline: 23642349]
4. Lin H, Lai C, Lin P, Friedman M, Salapatas AM, Chang H, et al. Clinical prediction model for obstructive sleep apnea among adult patients with habitual snoring. Otolaryngol Head Neck Surg 2019 Jul 02;161(1):178-185. [doi: 10.1177/0194599819839999] [Medline: 30935275]
5. 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]
6. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019 Jan 01;170(1):W1. [doi: 10.7326/m18-1377]
7. Ustun B, Westover MB, Rudin C, Bianchi MT. Clinical prediction models for sleep apnea: the importance of medical history over symptoms. J Clin Sleep Med 2016 Feb;12(2):161-168 [FREE Full text] [doi: 10.5664/jcsm.5476] [Medline: 26350602]
8. Haberfeld C, Sheta A, Hossain MS, Turabieh H, Surani S. SAS mobile application for diagnosis of obstructive sleep apnea utilizing machine learning models. In: Proceedings of the 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). 2020 Presented at: 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON); Oct 28-31, 2020; New York, NY, USA. [doi: 10.1109/uemcon51285.2020.9298041]
9. Kapuniai LE, Andrew DJ, Crowell DH, Pearce JW. Identifying sleep apnea from self-reports. Sleep 1988 Oct;11(5):430-436. [doi: 10.1093/sleep/11.5.430] [Medline: 3227223]
10. Morris LG, Kleinberger A, Lee KC, Liberatore LA, Burschtin O. Rapid risk stratification for obstructive sleep apnea, based on snoring severity and body mass index. Otolaryngol Head Neck Surg 2008 Nov 24;139(5):615-618. [doi: 10.1016/j.otohns.2008.08.026] [Medline: 18984252]
11. Deegan PC, McNicholas WT. Predictive value of clinical features for the obstructive sleep apnoea syndrome. Eur Respir J 1996 Jan;9(1):117-124 [FREE Full text] [doi: 10.1183/09031936.96.09010117] [Medline: 8834344]
12. Viner S, Szalai JP, Hoffstein V. Are history and physical examination a good screening test for sleep apnea? Ann Intern Med 1991 Sep 01;115(5):356-359. [doi: 10.7326/0003-4819-115-5-356] [Medline: 1863025]
13. Keenan SP, Anderson B, Wiggs B, Ryan CF, Fleetham JA. The predictive accuracy of home oximetry in patients with suspected obstructive sleep apnea. Sleep 1993 Dec;16(8 Suppl):S133-S134. [doi: 10.1093/sleep/16.suppl_8.s133] [Medline: 8178005]
14. Hoffstein V, Szalai JP. Predictive value of clinical features in diagnosing obstructive sleep apnea. Sleep 1993 Feb;16(2):118-122. [Medline: 8446830]
15. Flemons WW, Whitelaw WA, Brant R, Remmers JE. Likelihood ratios for a sleep apnea clinical prediction rule. Am J Respir Crit Care Med 1994 Nov;150(5 Pt 1):1279-1285. [doi: 10.1164/ajrccm.150.5.7952553] [Medline: 7952553]
16. Vaidya AM, Petruzzelli GJ, Walker RP, McGee D, Gopalsami C. Identifying obstructive sleep apnea in patients presenting for laser-assisted uvulopalatoplasty. Laryngoscope 1996 Apr;106(4):431-437. [doi: 10.1097/00005537-199604000-00008] [Medline: 8614217]
17. Pradhan PS, Gliklich RE, Winkelman J. Screening for obstructive sleep apnea in patients presenting for snoring surgery. Laryngoscope 1996 Nov;106(11):1393-1397. [doi: 10.1097/00005537-199611000-00016] [Medline: 8914907]
18. Friedman M, Tanyeri H, La Rosa M, Landsberg R, Vaidyanathan K, Pieri S, et al. Clinical predictors of obstructive sleep apnea. Laryngoscope 1999 Dec;109(12):1901-1907. [doi: 10.1097/00005537-199912000-00002] [Medline: 10591345]
19. Dixon JB, Schachter LM, O'Brien PE. Predicting sleep apnea and excessive day sleepiness in the severely obese: indicators for polysomnography. Chest 2003 Apr;123(4):1134-1141. [doi: 10.1378/chest.123.4.1134] [Medline: 12684304]
20. Martinez-Rivera C, Abad J, Fiz JA, Rios J, Morera J. Usefulness of truncal obesity indices as predictive factors for obstructive sleep apnea syndrome. Obesity (Silver Spring) 2008 Jan;16(1):113-118 [FREE Full text] [doi: 10.1038/oby.2007.20] [Medline: 18223622]

21. Herzog M, Kühnel T, Bremert T, Herzog B, Hosemann W, Kaftan H. The upper airway in sleep-disordered breathing: a clinical prediction model. Laryngoscope 2009 Apr;119(4):765-773. [doi: 10.1002/lary.20153] [Medline: 19266582]
22. Yeh P, Lee Y, Lee W, Chen S, Ho S, Peng W, et al. Clinical predictors of obstructive sleep apnea in Asian bariatric patients. Obes Surg 2010 Jan 12;20(1):30-35. [doi: 10.1007/s11695-009-9854-2] [Medline: 19434465]
23. Hukins C. Mallampati class is not useful in the clinical assessment of sleep clinic patients. J Clin Sleep Med 2010 Dec 15;06(06):545-549. [doi: 10.5664/jcsm.27987]
24. Musman S, Passos VM, Silva IB, Barreto SM. Evaluation of a prediction model for sleep apnea in patients submitted to polysomnography. J Bras Pneumol 2011 Feb;37(1):75-84 [FREE Full text] [doi: 10.1590/s1806-37132011000100012] [Medline: 21390435]
25. Sareli AE, Cantor CR, Williams NN, Korus G, Raper SE, Pien G, et al. Obstructive sleep apnea in patients undergoing bariatric surgery--a tertiary center experience. Obes Surg 2011 Mar 11;21(3):316-327. [doi: 10.1007/s11695-009-9928-1] [Medline: 19669842]
26. Tseng MH, Hsu HC, Chang CC, Ting H, Wu HC, Tang PH. Development of an intelligent app for obstructive sleep apnea prediction on android smartphone using data mining approach. In: Proceedings of the 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing. 2012 Presented at: 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing; Sep 04-07, 2012; Fukuoka, Japan. [doi: 10.1109/uic-atc.2012.89]
27. Sahin M, Bilgen C, Tasbakan MS, Midilli R, Basoglu OK. A clinical prediction formula for apnea-hypopnea index. Int J Otolaryngol 2014;2014:438376-438375 [FREE Full text] [doi: 10.1155/2014/438376] [Medline: 25349613]
28. Ting H, Mai Y, Hsu H, Wu H, Tseng M. Decision tree based diagnostic system for moderate to severe obstructive sleep apnea. J Med Syst 2014 Sep 11;38(9):94. [doi: 10.1007/s10916-014-0094-1] [Medline: 25012477]
29. Sutherland K, Lee RW, Petocz P, Chan TO, Ng S, Hui DS, et al. Craniofacial phenotyping for prediction of obstructive sleep apnoea in a Chinese population. Respirology 2016 Aug 15;21(6):1118-1125. [doi: 10.1111/resp.12792] [Medline: 27083503]
30. Del Brutto OH, Mera RM, Recalde BY, Castillo PR. Assessment of neck grasp as a screening tool for identifying obstructive sleep apnea in community-dwelling older adults. J Prim Care Community Health 2020 Dec 24;11:2150132720984421 [FREE Full text] [doi: 10.1177/2150132720984421] [Medline: 33356814]
31. Chen YF, Chen JH, Lin YJ, Tai CJ. Diagnosis and prediction of patients with severe obstructive apneas using support vector machine. In: Proceedings of the 2008 International Conference on Machine Learning and Cybernetics. 2008 Presented at: 2008 International Conference on Machine Learning and Cybernetics; Jul 12-15, 2008; Kunming. [doi: 10.1109/icmlc.2008.4620964]
32. Kirby SD, Eng P, Danter W, George CF, Francovic T, Ruby RR, et al. Neural network prediction of obstructive sleep apnea from clinical criteria. Chest 1999 Aug;116(2):409-415. [doi: 10.1378/chest.116.2.409] [Medline: 10453870]
33. Lam B, Ip MS, Tench E, Ryan CF. Craniofacial profile in Asian and white subjects with obstructive sleep apnoea. Thorax 2005 Jun 01;60(6):504-510 [FREE Full text] [doi: 10.1136/thx.2004.031591] [Medline: 15923252]
34. Julià-Serdà G, Pérez-Peñate G, Saavedra-Santana P, Ponce-González M, Valencia-Gallardo JM, Rodríguez-Delgado R, et al. Usefulness of cephalometry in sparing polysomnography of patients with suspected obstructive sleep apnea. Sleep Breath 2006 Dec 20;10(4):181-187. [doi: 10.1007/s11325-006-0073-y] [Medline: 17053929]
35. Polat K, Yosunkaya P, Güneş S. Comparison of different classifier algorithms on the automated detection of obstructive sleep apnea syndrome. J Med Syst 2008 Jun 16;32(3):243-250. [doi: 10.1007/s10916-008-9129-9] [Medline: 18444362]
36. Lee RW, Petocz P, Prvan T, Chan AS, Grunstein RR, Cistulli PA. Prediction of obstructive sleep apnea with craniofacial photographic analysis. Sleep 2009 Jan;32(1):46-52 [FREE Full text] [Medline: 19189778]
37. Rofail LM, Wong KK, Unger G, Marks GB, Grunstein RR. The utility of single-channel nasal airflow pressure transducer in the diagnosis of OSA at home. Sleep 2010 Aug;33(8):1097-1105 [FREE Full text] [doi: 10.1093/sleep/33.8.1097] [Medline: 20815193]
38. Chen N, Chen M, Li H, Chen C, Wang P. A two-tier screening model using quality-of-life measures and pulse oximetry to screen adults with sleep-disordered breathing. Sleep Breath 2011 Sep 7;15(3):447-454. [doi: 10.1007/s11325-010-0356-1] [Medline: 20449670]
39. Bucca C, Brussino L, Maule MM, Baldi I, Guida G, Culla B, et al. Clinical and functional prediction of moderate to severe obstructive sleep apnoea. Clin Respir J 2011 Oct;5(4):219-226. [doi: 10.1111/j.1752-699X.2010.00223.x] [Medline: 21801324]
40. Bouloukaki I, Kapsimalis F, Mermigkis C, Kryger M, Tzanakis N, Panagou P, et al. Prediction of obstructive sleep apnea syndrome in a large Greek population. Sleep Breath 2011 Dec 25;15(4):657-664. [doi: 10.1007/s11325-010-0416-6] [Medline: 20872180]
41. Sun LM, Chiu H, Chuang CY, Liu L. A prediction model based on an artificial intelligence system for moderate to severe obstructive sleep apnea. Sleep Breath 2011 Sep 4;15(3):317-323. [doi: 10.1007/s11325-010-0384-x] [Medline: 20602177]
42. Laporta R, Anandam A, El-Solh AA. Screening for obstructive sleep apnea in veterans with ischemic heart disease using a computer-based clinical decision-support system. Clin Res Cardiol 2012 Sep 3;101(9):737-744. [doi: 10.1007/s00392-012-0453-1] [Medline: 22476823]

43. Hang L, Lin H, Cheng C, Chiang JY, Wang H, Chen Y. Diagnosis of severe obstructive sleep apnea with model designed using genetic algorithm and ensemble support vector machine. Appl Math Inf Sci 2015 Feb 1;9(1L):149-157. [doi: 10.12785/amis/091l19]
44. Hang L, Wang H, Chen J, Hsu J, Lin H, Chung W, et al. Validation of overnight oximetry to diagnose patients with moderate to severe obstructive sleep apnea. BMC Pulm Med 2015 Mar 20;15(1):24 [FREE Full text] [doi: 10.1186/s12890-015-0017-z] [Medline: 25880649]
45. Bozkurt S, Bostanci A, Turhan M. Can statistical machine learning algorithms help for classification of obstructive sleep apnea severity to optimal utilization of polysomno graphy resources? Methods Inf Med 2018 Jan 24;56(04):308-318. [doi: 10.3414/me16-01-0084]
46. Ferreira-Santos D, Rodrigues PP. Improving diagnosis in obstructive sleep apnea with clinical data: a Bayesian network approach. In: Proceedings of the IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS). 2017 Presented at: IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS); Jun 22-24, 2017; Thessaloniki, Greece. [doi: 10.1109/cbms.2017.19]
47. Liu W, Wu H, Juang J, Wisniewski A, Lee H, Wu D, et al. Prediction of the severity of obstructive sleep apnea by anthropometric features via support vector machine. PLoS One 2017 May 4;12(5):e0176991 [FREE Full text] [doi: 10.1371/journal.pone.0176991] [Medline: 28472141]
48. Manoochehri Z, Rezaei M, Salari N, Khazaie H, Khaledi Paveh B, Manoochehri S. The prediction of obstructive sleep apnea using data mining approaches. Arch Iran Med 2018 Oct 01;21(10):460-465 [FREE Full text] [Medline: 30415554]
49. Manoochehri Z, Salari N, Rezaei M, Khazaie H, Manoochehri S, paveh B. Comparison of support vector machine based on genetic algorithm with logistic regression to diagnose obstructive sleep apnea. J Res Med Sci 2018;23(1):65. [doi: 10.4103/jrms.jrms_357_17]
50. Xu H, Zhao X, Shi Y, Li X, Qian Y, Zou J, et al. Development and validation of a simple-to-use clinical nomogram for predicting obstructive sleep apnea. BMC Pulm Med 2019 Jan 18;19(1):18 [FREE Full text] [doi: 10.1186/s12890-019-0782-1] [Medline: 30658615]
51. Ferreira-Santos D, Rodrigues PP. A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches. Int J Data Sci Anal 2018 Apr 13;8(4):339-349. [doi: 10.1007/s41060-018-0118-x]
52. Keshavarz Z, Rezaee R, Nasiri M, Pournik O. Obstructive sleep apnea: a prediction model using supervised machine learning method. Stud Health Technol Inform 2020 Jun 26;272:387-390. [doi: 10.3233/SHTI200576] [Medline: 32604683]
53. Chen W, Feng J, Wang Y, Wang C, Dong Z. Development and validation of a nomogram for predicting obstructive sleep apnea in bariatric surgery candidates. Nat Sci Sleep 2021 Jun;13:1013-1023. [doi: 10.2147/nss.s316674]
54. Hsu Y, Wang J, Huang P, Chien Y, Chiu C, Lin C. Integrating domain knowledge with machine learning to detect obstructive sleep apnea: snore as a significant bio-feature. J Sleep Res 2022 Apr 21;31(2):e13487. [doi: 10.1111/jsr.13487] [Medline: 34549473]
55. Crocker BD, Olson LG, Saunders NA, Hensley MJ, McKeon JL, Allen KM, et al. Estimation of the probability of disturbed breathing during sleep before a sleep study. Am Rev Respir Dis 1990 Jul;142(1):14-18. [doi: 10.1164/ajrccm/142.1.14] [Medline: 2368960]
56. Pillar G, Peled N, Katz N, Lavie P. Predictive value of specific risk factors, symptoms and signs, in diagnosing obstructive sleep apnoea and its severity. J Sleep Res 1994 Dec;3(4):241-244 [FREE Full text] [doi: 10.1111/j.1365-2869.1994.tb00137.x] [Medline: 10607131]
57. Maislin G, Pack AI, Kribbs NB, Smith PL, Schwartz AR, Kline LR, et al. A survey screen for prediction of apnea. Sleep 1995 Apr;18(3):158-166. [doi: 10.1093/sleep/18.3.158] [Medline: 7610311]
58. Kushida CA, Efron B, Guilleminault C. A predictive morphometric model for the obstructive sleep apnea syndrome. Ann Intern Med 1997 Oct 15;127(8 Pt 1):581-587. [doi: 10.7326/0003-4819-127-8_part_1-199710150-00001] [Medline: 9341055]
59. El-Solh AA, Mador MJ, Ten-Brock E, Shucard DW, Abul-Khoudoud M, Grant BJ. Validity of neural network in sleep apnea. Sleep 1999 Feb 01;22(1):105-111. [doi: 10.1093/sleep/22.1.105] [Medline: 9989371]
60. Zerah-Lancner F, Lofaso F, d'Ortho MP, Delclaux C, Goldenberg F, Coste A, et al. Predictive value of pulmonary function parameters for sleep apnea syndrome. Am J Respir Crit Care Med 2000 Dec;162(6):2208-2212. [doi: 10.1164/ajrccm.162.6.2002002] [Medline: 11112139]
61. Rodsutti J, Hensley M, Thakkinstian A, D'Este C, Attia J. A clinical decision rule to prioritize polysomnography in patients with suspected sleep apnea. Sleep 2004 Jun 15;27(4):694-699. [doi: 10.1093/sleep/27.4.694] [Medline: 15283004]
62. Khoo S, Poh H, Chan Y, Ngerng W, Shi D, Lim TK. Diagnostic characteristics of clinical prediction models for obstructive sleep apnea in different clinic populations. Sleep Breath 2011 Sep 4;15(3):431-437. [doi: 10.1007/s11325-010-0354-3] [Medline: 20440569]
63. Zou J, Guan J, Yi H, Meng L, Xiong Y, Tang X, et al. An effective model for screening obstructive sleep apnea: a large-scale diagnostic study. PLoS One 2013 Dec 2;8(12):e80704 [FREE Full text] [doi: 10.1371/journal.pone.0080704] [Medline: 24312494]

64. Karamanli H, Yalcinoz T, Yalcinoz MA, Yalcinoz T. A prediction model based on artificial neural networks for the diagnosis of obstructive sleep apnea. Sleep Breath 2016 May 19;20(2):509-514. [doi: 10.1007/s11325-015-1218-7] [Medline: 26087718]
65. Tawaranurak K, Kamolphiwong S, Sae-Wong S, Vasupongayya S, Kamolphiwong T, Bumrungsena C, et al. Validity of a new prediction model to identify patients at risk for obstructive sleep apnea hypopnea syndrome. Ear Nose Throat J 2021 Jan 04:145561320986045 [FREE Full text] [doi: 10.1177/0145561320986045] [Medline: 33393817]
66. Park D, Kim J, Park B, Kim HJ. Risk factors and clinical prediction formula for the evaluation of obstructive sleep apnea in Asian adults. PLoS One 2021 Feb 2;16(2):e0246399 [FREE Full text] [doi: 10.1371/journal.pone.0246399] [Medline: 33529265]
67. Rowley JA, Aboussouan LS, Badr MS. The use of clinical prediction formulas in the evaluation of obstructive sleep apnea. Sleep 2000 Nov 01;23(7):929-938. [doi: 10.1093/sleep/23.7.929] [Medline: 11083602]
68. Soares MC, de Azeredo Bittencourt LR, Zonato AI, Gregório LC. Application of the Kushida morphometric model in patients with sleep-disordered breathing. Brazilian J Otorhinolaryngol 2006 Jul;72(4):541-548. [doi: 10.1016/s1808-8694(15)31002-8]

## Abbreviations

**AASM:** American Academy of Sleep Medicine
**AHI:** apnea-hypopnea index
**AUC:** area under the receiver operating characteristic curve
**NC:** neck circumference
**OSA:** obstructive sleep apnea
**PROBAST:** Prediction Model Risk of Bias Assessment Tool
**PROSPERO:** International Prospective Register of Systematic Reviews
**PSG:** polysomnography

*Edited by R Kukafka; submitted 10.05.22; peer-reviewed by M Pičulin, R Damaševičius; comments to author 13.06.22; revised version received 20.06.22; accepted 19.07.22; published 30.09.22.*

*Please cite as:*
*Ferreira-Santos D, Amorim P, Silva Martins T, Monteiro-Soares M, Pereira Rodrigues P*
*Enabling Early Obstructive Sleep Apnea Diagnosis With Machine Learning: Systematic Review*
*J Med Internet Res 2022;24(9):e39452*
*URL: https://www.jmir.org/2022/9/e39452*
*doi:10.2196/39452*
*PMID:36178720*

Viewpoint

# A Revised Hippocratic Oath for the Era of Digital Health

Bertalan Meskó[1], MD, PhD; Brennan Spiegel[2], MD, PhD

[1]The Medical Futurist Institute, Budapest, Hungary

[2]Cedars-Sinai Health System, Los Angeles, CA, United States

**Corresponding Author:**
Bertalan Meskó, MD, PhD
The Medical Futurist Institute
Povl Bang-Jensen u 2/B1 4/1
Budapest, 1118
Hungary
Phone: 36 703807260
Email: berci@medicalfuturist.com

## Abstract

Physicians have been taking the Hippocratic Oath for centuries. The Oath contains a set of ethical rules designed to guide physicians through their profession; it articulates a set of true north principles that govern the practice of medicine. The Hippocratic Oath has undergone several revisions, most notably in 1948 by the World Medical Association. However, in an era of rapid change in medicine, we believe it is time to update the Oath with modest but meaningful additions so that it optimally reflects 21st century health care. The rise of digital health has dramatically changed the practice of medicine in a way that could not have been easily predicted at the time Hippocrates outlined his ethical principles of medicine. Digital health is a broad term that encompasses use of digital devices and platforms, including electronic health records, patient-provider portals, mobile health apps, wearable biosensors, artificial intelligence, social media platforms, and medical extended reality, to improve the process and outcomes of health care delivery. These technologies have driven a cultural transformation in the delivery of care. We offer modest suggestions to help prompt discussion and contemplation about the current Oath and its relevancy to our changing times. Our suggestions are not meant to be a definitive set of final recommendations. Rather, we propose new text that bodies such as the World Medical Association might consider integrating into an updated Oath, just as previous changes were adopted to ensure the Oath remains relevant and impactful for all physicians and their patients.

***(J Med Internet Res 2022;24(9):e39177)*** doi:10.2196/39177

**KEYWORDS**

hippocratic oath; digital health; eHealth; future; automation; ethics; viewpoint; medical perspective; physician perspective; ethical; digital divide; artificial intelligence; moral

## A Brief History of the Hippocratic Oath

Physicians have been taking the Hippocratic Oath for centuries [1]. The Oath contains a set of ethical rules designed to guide physicians through their profession; it articulates a set of true north principles that govern the practice of medicine.

Although there are no explicit penalties enforced upon physicians who break the Oath, adherence with its principles remains a time-honored tradition by practicing clinicians. The original Hippocratic Oath describes ideals that are timely and relevant even in the 21st century: to treat patients to the best of one's ability; to preserve a patient's privacy; and to faithfully teach the art of medicine to the next generation [2].

Most physicians believe the Oath still has relevance today, although viewpoints remain varied and there are few empirical studies that have formally evaluated sentiment about the Oath. In a non–peer-reviewed survey conducted in 2016, Medscape reported opinions about the modern relevance of Hippocrates' pledge [3]. Nearly 3000 physicians and medical students responded. Reactions were markedly different, particularly when stratified by age. Of those under age 34 years, only 39% said the Oath was still meaningful, whereas 70% of respondents aged 65 years and older positively endorsed the pledge. Despite these varying views about the Oath, coupled with limited peer-reviewed data on this topic, most medical schools still ask their students to recite either the classic or modified form of the Oath.

The Hippocratic Oath has undergone several revisions, most notably in 1948 by the World Medical Association; this version is called the Declaration of Geneva [4]. Changes included removing lines such as "My colleagues will be my sisters and brothers" and added lines such as "I will respect the autonomy and dignity of my patient." However, in an era of rapid change in medicine, we believe it is time to update the Oath with modest but meaningful additions so that it optimally reflects 21st century health care.

## *Why a Revised Hippocratic Oath Is Warranted in the Era of Digital Health*

The rise of digital health has dramatically changed the practice of medicine in a way that could not have been easily predicted at the time Hippocrates outlined his ethical principles of medicine. Digital health is a broad term that encompasses use of digital devices and platforms, including electronic health records, patient-provider portals, mobile health apps, wearable biosensors, artificial intelligence (AI), social media platforms, and medical extended reality, to improve the process and outcomes of health care delivery. These technologies have driven a cultural transformation in the delivery of care [5].

Remote patient monitoring, for example, affords patients and doctors a more complete and accurate picture of disease progression outside the walls of a hospital, clinic, or research center. The data from mobile technologies can now be shared between patient and provider, allowing greater collaboration, stronger therapeutic partnerships, enhanced shared decision-making, and an increasing shift to preventive and proactive care in lieu of reactive care.

Extended reality technologies, such as virtual reality and augmented reality, provide opportunities to go beyond the traditional exam room and introduce new ways of blending behavioral and psychosocial care with traditional biomedical care [6]. AI is massively expanding our ability to diagnose and treat patients, but simultaneously raising significant ethical debates about the potential for misuse of powerful and potentially biased algorithms [7]. Social media platforms have become a digital town hall for all manner of health care information exchange, further democratizing access to information historically under the sole purview of physicians.

In short, the cultural transformation enabled by digital health is rapidly changing the practice of medicine from a tradition of physician-driven decisions based on limited, institutionally owned data, to shared decision-making based on expansive data across platforms, owned and shared by patients, that reflects biopsychosocial well-being across broad disease trajectories and illness experiences along a range of geographies, demographics, and sociocultural communities. In light of this transformation, we believe it is warranted to modestly update the Hippocratic Oath so it optimally reflects 21st century medicine.

## *Proposed Revisions to the Hippocratic Oath*

In the context of changes produced by advances in digital health, we suggest that the following principles should be reflected within a modernized Hippocratic Oath, presented in order of the current Oath's text.

### Recognize a Broader Origin of Scientific Gains in Medicine

The Hippocratic Oath entreats physicians to "respect the hard-won scientific gains of those physicians on whose steps I walk." This statement suggests that research is conducted by physicians only, when in fact medical advances originate from many stakeholders beyond physicians. In the era of digital health and democratized care, research arises not only from physicians and nonclinical researchers, but also from patients who both contribute their own data and meaningfully participate in research through patient-centered models such as those supported by the Patient-Centered Outcomes Research Institute [8]. We suggest the following edits, shown in italics:

- I will respect the hard-won scientific gains of those physicians, *researchers, and patients* in whose steps I walk, and gladly share such knowledge as is mine with those who are to follow.

### Acknowledge Both "Sick Care" and Preventive "Health Care"

The Oath focuses on treating the sick but is silent on the role of preventive medicine for the well, yet modern medicine emphasizes the importance of preventive care across physical, mental, and social realms of health, not just reactive "sick care" for the ill. Advances in digital health place an emphasis on predictive analytics using remotely collected data, and precision medicine aims to identify early signs of disease to inform timely preventive care. Given these forces that are shifting medical delivery from reactive "sick care" to preventive "health care," we propose the following simple addition to the Oath, shown in italics:

- I will apply, for the benefit of *the healthy and* the sick, all measures [that] are required, avoiding those twin traps of overtreatment and therapeutic nihilism.

### Reflect the Intrinsic Use of Digital Technology in the Practice of Medicine

Considering the massive advances in technology, coupled with the reality that technology now plays an everyday role in the delivery of health care, we believe the Hippocratic Oath should reflect the foundational role of digital health in patient care. As advances in AI, robotics, virtual reality, mobile health apps, wearable biosensors, and portable diagnostic devices continue to expand, we believe the Oath should acknowledge the growing and permanent impact these technologies now exert on care delivery, as follows:

- I will remember that there is an art to medicine as well as science, and that warmth, sympathy, and understanding may outweigh the surgeon's knife, the chemist's drug, *or the programmer's algorithm.*

### Validate the Patient Role in an Equal-Level Partnership With Their Physician

The Hippocratic Oath encourages physicians to say, "I know not" when they are unsure how to treat a patient, and to "call in my colleagues when the skills of another are needed for a patient's recovery." These are laudable sentiments that any clinician should follow. However, the Oath should ideally acknowledge that patients, too, can help with diagnosis and treatment.

Although the tradition of medicine reinforces a hierarchical patient-doctor relationship driven by information asymmetry, patients now have expansive access to credible information about biomedical sciences, increasingly generate their own biometric health data through wearable biosensors, and monitor their own psychometric scores through apps; these data sources are now part of clinical practice.

Although physicians have more experience prescribing treatments and monitoring a wide range of diseases than nonphysician patients, the Oath should recognize that patients are the experts of their personal illness experience. When engaged collaboratively by their physicians, patients can deliver meaningful insights that shape diagnostic and care plans. In light of this paradigm shift that has accelerated in the era of digital health, we suggest the following italicized additions to the Oath:

- *I will treat my patients in an equal-level partnership, and* I will not be ashamed to say 'I know not,' nor will I fail to call in my colleagues when the skills of another are needed for a patient's recovery.

### Address Data Privacy

Respecting patients' privacy is a primary passage in the Oath. However, the concept of privacy now extends beyond safekeeping conversations to guarding the "big data" generated in the care of every patient in modern health care. We suggest the following addition:

- I will respect the privacy of my patients *and their data*, for their problems are not disclosed to me that the world may know.

### Emphasize the Primacy of Treating Patients, Not Their Data

The explosion in big data surrounding health care is transforming how doctors care for and interact with their patients. AI in particular has a vast potential to automate processes in health care and potentially overtake certain roles and responsibilities normally filled by clinicians. Nonetheless, physician must always remain focused on their patients, including their personal stories and their biopsychosocial well-being beyond their digital fingerprints and big data analytics. AI will never replace medical professionals, although physicians who embrace AI may eventually replace those who do not. We propose the following additions to the Oath to reflect these considerations:

- I will remember that I do not treat a fever chart, a cancerous growth, *a data point, or an algorithm's suggestion*, but a human being.

## Conclusions

The Hippocratic Oath remains an important pledge that modern physicians should continue to honor. The Oath outlines principles that remain relevant in the 21st century. However, advances in digital health science and technology have catalyzed a cultural revolution in the delivery of health care.

We believe that it is now justified to modify the Hippocratic Oath—even if modestly—to reflect the digital health revolution, advances in patient empowerment, and the evolving role of technology in the everyday practice of medicine.

We offer these modest suggestions to help prompt discussion and contemplation about the current Oath and its relevancy to our changing times. Our suggestions are not meant to be a definitive set of final recommendations. Rather, we propose new text that bodies such as the World Medical Association might consider integrating into an updated Oath, just as previous changes were adopted to ensure the Oath remains relevant and impactful for all physicians and their patients (Multimedia Appendix 1).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Revised Hippocratic Oath.
[DOCX File , 14 KB - jmir_v24i9e39177_app1.docx ]

## References

1. Miles SH. The Hippocratic Oath and the Ethics of Medicine. Oxford: Oxford University Press; 2005.
2. Gakis D. The Hippocratic Oath today. MOJS 2016 May 31;3(2):35. [doi: 10.15406/mojs.2016.03.00039]
3. Frellick M. Youngest, Oldest Physicians Diverge on Hippocratic Oath. Medscape. 2017. URL: https://www.medscape.com/viewarticle/880688 [accessed 2022-08-24]
4. Parsa-Parsi RW. The Revised Declaration of Geneva: A Modern-Day Physician's Pledge. JAMA 2017 Nov 28;318(20):1971-1972. [doi: 10.1001/jama.2017.16230] [Medline: 29049507]
5. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]

6. Spiegel B. VRx: How Virtual Therapeutics Will Revolutionize Medicine. New York: Basic Books; 2020.
7. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York: Basic Books; 2019.
8. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/ [accessed 2022-08-24]

## Abbreviations

**AI:** artificial intelligence

---

*Edited by T Leung, G Eysenbach; submitted 02.05.22; peer-reviewed by C Nebecker, R Hendricks-Sturrup, R Rastmanesh, J Kaswija; comments to author 01.06.22; revised version received 28.07.22; accepted 19.08.22; published 07.09.22.*

*Please cite as:*
*Meskó B, Spiegel B*
*A Revised Hippocratic Oath for the Era of Digital Health*
*J Med Internet Res 2022;24(9):e39177*
*URL: https://www.jmir.org/2022/9/e39177*
*doi:10.2196/39177*
*PMID:36069845*

Viewpoint

# Dynamic Digital Twin: Diagnosis, Treatment, Prediction, and Prevention of Disease During the Life Course

Skander Tahar Mulder[1*], PhD; Amir-Houshang Omidvari[2,3*], MD, PhD; Anja J Rueten-Budde[4], PhD; Pei-Hua Huang[5], PhD; Ki-Hun Kim[6], PhD; Babette Bais[7], PhD; Melek Rousian[7], MD, PhD; Rihan Hai[8], PhD; Can Akgun[8,9], PhD; Jeanine Roeters van Lennep[10], MD, PhD; Sten Willemsen[4], PhD; Peter R Rijnbeek[11], PhD; David MJ Tax[1], PhD; Marcel Reinders[1], IR, PhD; Eric Boersma[2], PhD; Dimitris Rizopoulos[4], PhD; Valentijn Visch[12], PhD; Régine Steegers-Theunissen[7*], MD, PhD

[1]Pattern Recognition Lab, Mathematics and Computer Science, Technical University Delft, Delft, Netherlands
[2]Department of Cardiology, Erasmus Medical Center, Rotterdam, Netherlands
[3]Department of Public Health, Erasmus Medical Center, Rotterdam, Netherlands
[4]Department of Biostatistics, Erasmus Medical Center, Rotterdam, Netherlands
[5]Department of Medical Ethics and Philosophy, Erasmus Medical Center, Rotterdam, Netherlands
[6]Department of Industrial Engineering, Pusan National University, Busan, Republic of Korea
[7]Obstetrics and Gynaecology, Erasmus Medical Center, Rotterdam, Netherlands
[8]Web Information Systems Group, Mathematics and Computer Science, Technical University of Delft, Delft, Netherlands
[9]Bioelectronics Section, Department of Microelectronics, Faculty of Electrical Engineering, Technical University Delft, Delft, Netherlands
[10]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, Netherlands
[11]Department of Medical Informatics, Erasmus Medical Center, Rotterdam, Netherlands
[12]Industrial Design, Mathematics and Computer Science, Technical University Delft, Delft, Netherlands
[*]these authors contributed equally

**Corresponding Author:**
Régine Steegers-Theunissen, MD, PhD
Obstetrics and Gynaecology
Erasmus Medical Center
Dr. Molewaterplein 40
Rotterdam, 3015GD
Netherlands
Phone: 31 10 7038256
Email: r.steegers@erasmusmc.nl

## *Abstract*

A digital twin (DT), originally defined as a virtual representation of a physical asset, system, or process, is a new concept in health care. A DT in health care is not a single technology but a domain-adapted multimodal modeling approach incorporating the acquisition, management, analysis, prediction, and interpretation of data, aiming to improve medical decision-making. However, there are many challenges and barriers that must be overcome before a DT can be used in health care. In this viewpoint paper, we build on the current literature, address these challenges, and describe a dynamic DT in health care for optimizing individual patient health care journeys, specifically for women at risk for cardiovascular complications in the preconception and pregnancy periods and across the life course. We describe how we can commit multiple domains to developing this DT. With our cross-domain definition of the DT, we aim to define future goals, trade-offs, and methods that will guide the development of the dynamic DT and implementation strategies in health care.

***(J Med Internet Res 2022;24(9):e35675)*** doi:10.2196/35675

**KEYWORDS**

digital health; digital twin; machine learning; artifical intelligence; obstetrics; cardiovascular; disease; health

## Background

### Overview of the Concept

Interest has been growing worldwide in the virtual representation of a physical asset, process, or system to model and simulate a real-world event. This representation, called a digital twin (DT), can represent the real-time performance or failure incidence of a deterministic system (eg, a factory production line) [1]. Since the introduction of the concept of a DT in 2003, DTs have been developed and used in areas such as construction, power, and oil and gas industries. In the aforementioned industries, the DT served mainly as an umbrella term for managing data and models of a closed system, and these models then guided actions taken in the system. However, for health care, DT is a new concept in need of a working definition. The DT in health care is not a single technology but a domain-adapted multimodal modeling approach incorporating methodologies for the acquisition, management, analysis, prediction, and interpretation of health-related data, aiming to improve medical decision-making and patient lifestyle choices.

Health care providers strive to obtain and use all relevant information on patients for personalized decision-making in clinical practice, considering the available evidence, clinical guidelines, and patient preferences. For example, wearable technologies such as smartwatches have enabled individuals to record their health data continuously, which can form a part of personal health records [2]. They can be very useful for clinical decision-making in practice, and they are currently in use for detection or monitoring of some disorders like atrial fibrillation; however, they are not yet completely in use in practice [3]. Unsolved challenges such as human information overload [4]; variable quality of routinely collected data from medical, lifestyle, mental, societal, and environmental sources; and limited interoperability of digital systems in health care are barriers to use these scattered large data sets, also referred to as Big Data. In addition, the current methodological approaches in evidence-based medicine are not able to use all this information for medical decision-making, as the population generating these data is heterogeneous, and previously discovered relationships between predictor and outcomes might not always hold for subpopulations. To alleviate these limitations, the underlying methods and workflow for data use need to be adapted. Furthermore, currently available analytical models in health care, such as decision aids using risk prediction models (eg, the U-prevent software) [5,6], are often developed based on limited data and the defined outcomes of interest of a small number of health care professional. Combining high-quality, scattered data sets for inference by applying a comprehensive multimodal data management approach in health care such as DT is deemed necessary to design applications that allow for diagnosis, treatment, prediction, and prevention of disease. Moreover, successfully implemented DT has the potential to improve health care by optimizing individual health outcomes from the earliest moment in the life course by offering personalized medicine in primary, secondary, and tertiary health care [7].

Since 2015, publication of papers describing the concept of applying a DT to health care to solve health challenges (eg, reducing adverse outcomes in certain patient populations or understanding important factors such as dietary factors or biomarkers) [8-13] has increased. However, literature descriptions indicate open theoretical and practical challenges that need to be addressed before we can work toward efficiently implementing a DT in health care. Challenges are related to (1) redefining a target population and matching data set to develop a DT or (2) implementing a physical or data-driven approach with limited data and follow-up to learn causal personal patterns. In addition, we suggest a convergence of experts in the medical, technical, scientific, and ethical domains, which is required to design a DT that meets requirements from several theoretical backgrounds. In our vision, DT development starts with the identification and description of a medical problem such as the prediction of cardiovascular complications in a patient who experienced hypertension during periconception (Figure 1). Solving such a medical problem involves striving for the best health- and value-based outcomes and addressing ethical value goals such as health equality (Figure 2). With these outcomes and value goals in place, a technological system incorporating best methodological practices such as reliability and reproducibility can be developed. After the initial system is developed, many feedback loops between domains occur in order to optimize requirements across domains. This digital health system is then studied to develop best practices using the scientific domain and keeping the healthy patient life course journey in mind. In an open system such as the life course, as opposed to a closed physical system, causal drivers for change in health conditions might not be known; therefore, the DT uses algorithms that learn drivers of dynamic data such as user experiences, medical data fields, medical scans, etc. To serve that purpose, the DT acquires time-series data and updates predictions in an online and data-driven manner. Additional challenges are defining medical applications of a DT in health care and considering ethical values.

In this viewpoint paper, we address the aforementioned challenges and put forward a vision for a dynamic DT in health care for optimizing individual patient health journeys characterized by healthy outcomes and positive experiences. Our cross-domain, which has the ethical, medical, technical, and scientific definition of the DT, will define future goals, trade-offs, and methods to guide the implementation strategies and iterative development of a dynamic DT in health care. In our paper, we propose a dynamic DT for health care that applies to the management of dynamic patient data and models. These models of health and disease are dynamic because they are developed, trained, tested, and updated to meet the dynamic value goals stemming from ethical, medical, and technological domains. Additionally, the effect of these interventions (such as personal lifestyle advice) guided by models is scientifically evaluated and updated to continually strive for better health outcomes (Figure 2).

**Figure 1.** Digital twin in clinical practice is modeled for hypertension starting from the periconception period until later stages of life. There are four discrete steps in which the digital twin can bring additional value by using patient data to recommend interventions optimized for the relevant values and outcomes of interest. CVD: cardiovascular disease; DT: digital twin.

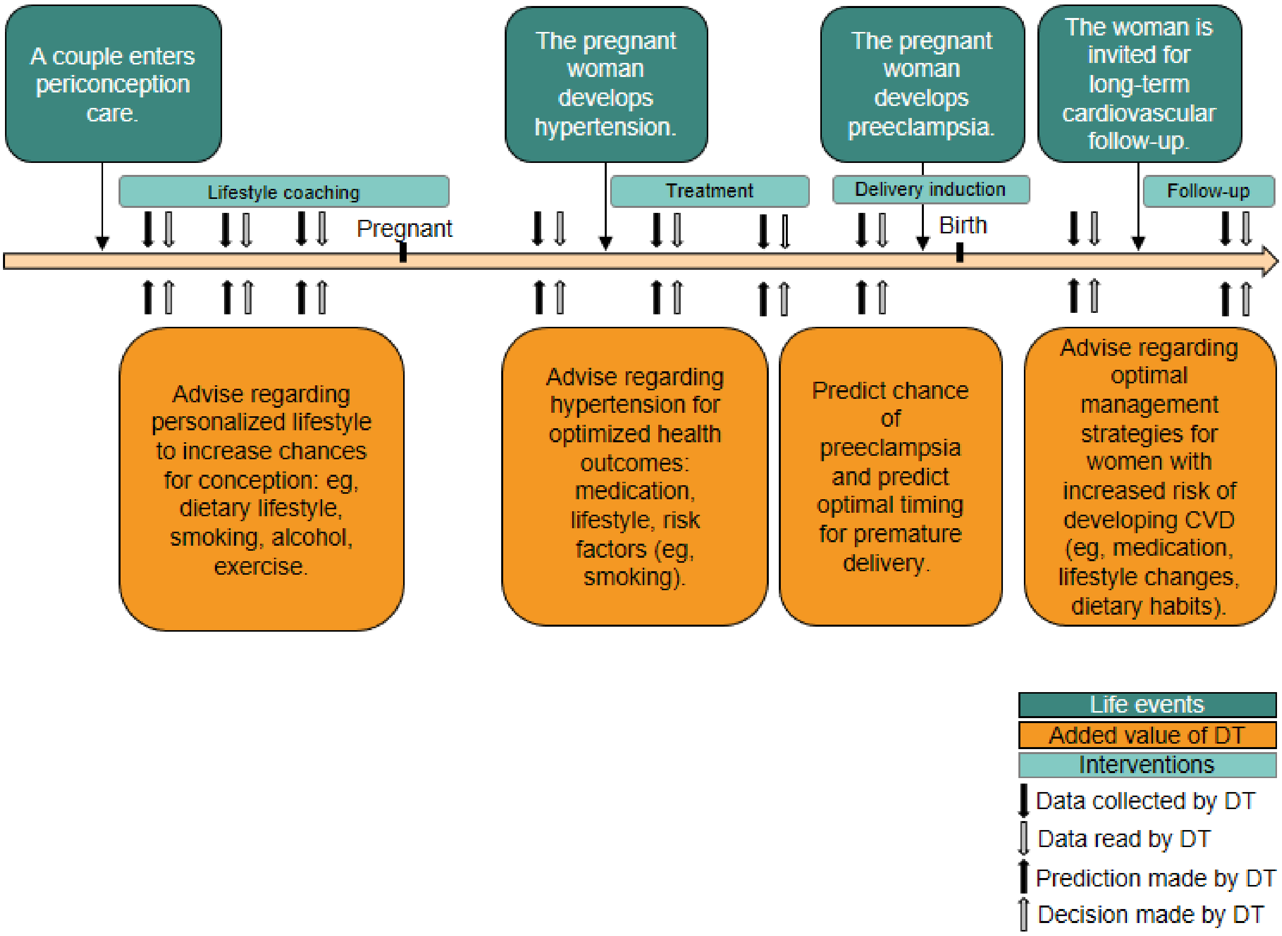
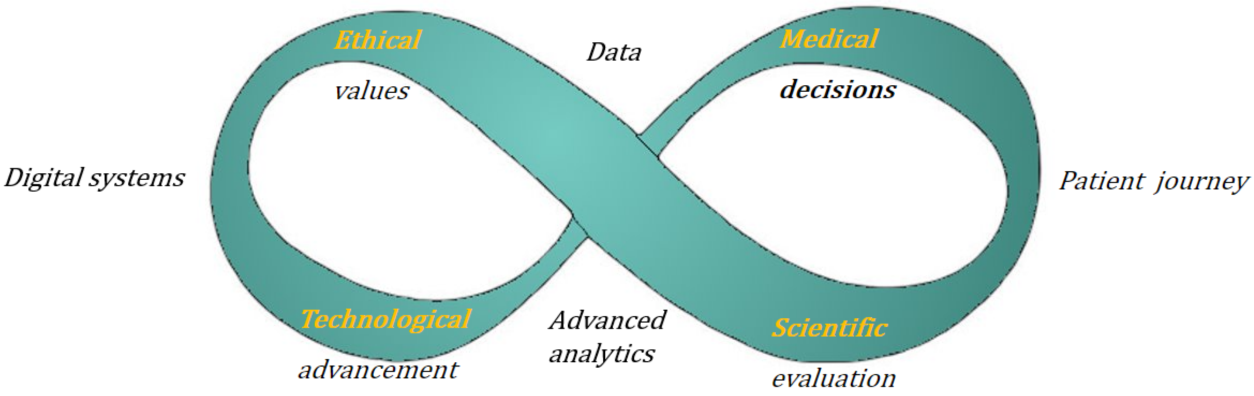


**Figure 2.** Digital systems and the patient health journey are improved by a continuous feedback loop across domains interacting to develop a digital twin.



## Potential Future Application of a DT in Health Care

Currently, the translation from data to evidence relating to disease ontology, causation, and effectiveness of treatments into clinical practice is a slow and partially data-driven process [14], and even for some of the recommendations in the clinical practice guidelines, there is no robust evidence available [15]. The dynamic DT can support a health journey by providing easy access to comprehensive patient data for the patient and health care providers. The DT can integrate data from health care sources like hospitals, general practitioners, and laboratories and from home health monitoring devices such as wearables.

Real-time integrated access to data will allow analysis on dynamically acquired data. This feature enables health care professionals to get a relevant view of factors influencing disease and health conditions of a patient, which can guide actions to optimize health care pathways. Different prediction algorithms can be used to answer medical questions, and real-world personalized predictions can be updated to meet the dynamic circumstances of individuals and their medical history. Testing these prediction models in the real world prospectively can help to define and research hypotheses about potential causal relationships between prediction algorithms and outcomes [16]. Formal testing and evaluating medical decisions in a DT can result in changing the focus of scientific studies to incorporate testing the implementation of a DT-powered intervention such as patient-tailored lifestyle advice in the absence of a data-driven decision framework. Therefore, this DT approach can expedite the process of knowledge translation into medical decision-making in clinical practice while striving for iterative improvement of elements in different domains of the DT.

## Example of a Patient Journey From the Earliest Moment of the Life Course

The periconceptional period refers to the 6 months around conception; it is the earliest and one of the most critical periods in life, with long-lasting impact on health and disease later in life and in future generations [17]. The patient journey and data generated during this period can help to illustrate our vision of the dynamic DT and how it would intervene in medical practice. To develop a DT from the earliest moment of the life course, the first set of data would include static (demographics) and dynamic (conditions such as subfertility, hypertension, lifestyle, and vital information from wearables) covariates relating to the parents. This DT supports dynamic medical decision-making (Figure 1). The first example is based on subfertility, for which enrollment in the evidence-based lifestyle coaching program is recommended. This eHealth coaching is not static but dynamic as it gives advice based on the variable input of the individual. For example, if healthy food intake is sufficient, parents will be empowered to maintain this behavior. A second example of how the DT can be used is aiding physicians to prevent short-term adverse outcomes of preeclampsia in pregnant women. The mobile phone app, a smartwatch, and lab measurements combined can yield predictions of risk before and after intervention, empowering medical decision-making and resulting in a better outcome for mother and baby. A third example to highlight the long-term benefits of a DT is hypertension. By mechanisms not fully understood, high blood pressure and preeclampsia during pregnancy yield increased cardiovascular risk for the unborn child and mother in later life years [18-20]. This data-driven association can make more comprehensive follow-up of these individuals valuable as this information can be used to monitor and potentially prevent worse outcomes such as cardiovascular disease. In our example, the DT has enacted medical decision rules driven by pattern recognition in patient data. In general terms, the rules obtained from pattern recognition systems are chosen so that they optimize for healthier outcomes.

## Why Now: Innovations That Power the DT

In 2016, a pivotal paper was published that described how data should be managed according to the following 4 principles: findable, accessible, interoperable, and reusable (FAIR) [21]. These principles promote data accessibility to power innovation and are the basis for the following 4 technical innovations that accelerate the progression of DT in health care. The first innovation is data storage, where we can securely store big data in the cloud by designing access rights for each service and algorithm [22]. Data collection innovations such as wearables, which can act as continuous vital sign monitoring systems, feed the new data to the storage keeping the data up to date and relevant. Connectivity innovations such as the Internet of Things then trigger the training and prediction algorithms with the updated information. Finally, computing innovations such as more powerful data processors like tensor processing units can fit bigger and more flexible models with higher performance to yield better representations of disease. These innovations have changed the way we form hypotheses about the physical reality. Increased data and processing power formats require and allow for new dynamic pattern recognition methods that can be described as computer pattern recognition or artificial intelligence [23].

These innovations power the DTs and allow for deep personalized predictions that leverage patient-specific dynamics such a specific disease pathway relating to biomarker panels based on genomics, metabolomics, or proteomics, which have previously been prioritized or identified. These predictions need to be translated in a concise way to an individual, patient or health care provider. For example, a score, dashboard, or written advice can guide individuals to healthy behavior. In our example, continuously monitored medical and lifestyle data of a couple contemplating pregnancy supplies the DT with data that can be used to give personalized integrated medical advice and brings personalized medicine closer [24].

## Themes in Pattern Recognition

There are multiple paradigms for the underlying data analytic platform of a DT relating to pattern recognition approaches. We can use a data-driven or physics-based approach. Physics-based models are based on the understanding of the phenomenon and formulation in a mathematical model, with underlying assumptions that potentially oversimplify the phenomenon [25]. Data-driven (associative) models avoid these assumptions, but they lack interpretability and are sensitive to bias in the model development data. We should choose our (inductive) bias in a way that allows us to optimize interpretability and predictive performance by leveraging domain medical expert–level knowledge. When the physical drivers of disease are not known, these associative data-driven methods can generally be used to diagnose patients, but they cannot identify causation. An alternative approach within the associative methodology is to enrich models with counterfactuals, which leverage counterfactual information to causally weight predictions [26].

In the DT for health care, pattern recognition algorithms are used for predicting continuous outcomes such as blood pressure, classification of diseases, and risk assessment. Medical data generally comprise different types of data and algorithms. For example, they include longitudinal data with missing predictors and variables, which can be used to enrich the model. Research on how to associatively predict with longitudinal data, population-level data, and data with missing values is ongoing. A statistical modeling approach, which can relate a limited number of predictors to a longitudinal outcome with missing values, uses linear mixed models [27]. Deep-learning methods [28,29], where we flexibly fit either a neural network architecture or many random kernels, are used when a large number of predictors is present. This modeling approach seems to perform well in this domain, but least absolute shrinkage and selection operator regression-based techniques may also yield good performance on this type of data [30].

Medical patient data can comprise higher dimensional data, such as echography, and multiple sensor data, such as electrocardiograms, which are continuously collected in a setting such as the coronary care unit. Imaging data is typically only acquired once and not collected continuously. Although each modality may be information rich, the number of measurements per individual may be limited, causing models to overfit on that individual. To prevent overfitting, we use feature selection and methods that can find a lower representation of these data, such as an auto encoder. Feature selection and engineering can prevent overfitting and summarize high-dimensional data in a feature vector, which can be used to describe disease progression.

As models are implemented in clinical practice, other challenges, such as model drift and the appearance of new class definitions, can arise that cause previously found associations to change. The online learning paradigm can adapt to this change by continuously updating the model with new information and thereby learning to adapt to changes in the environment [31].

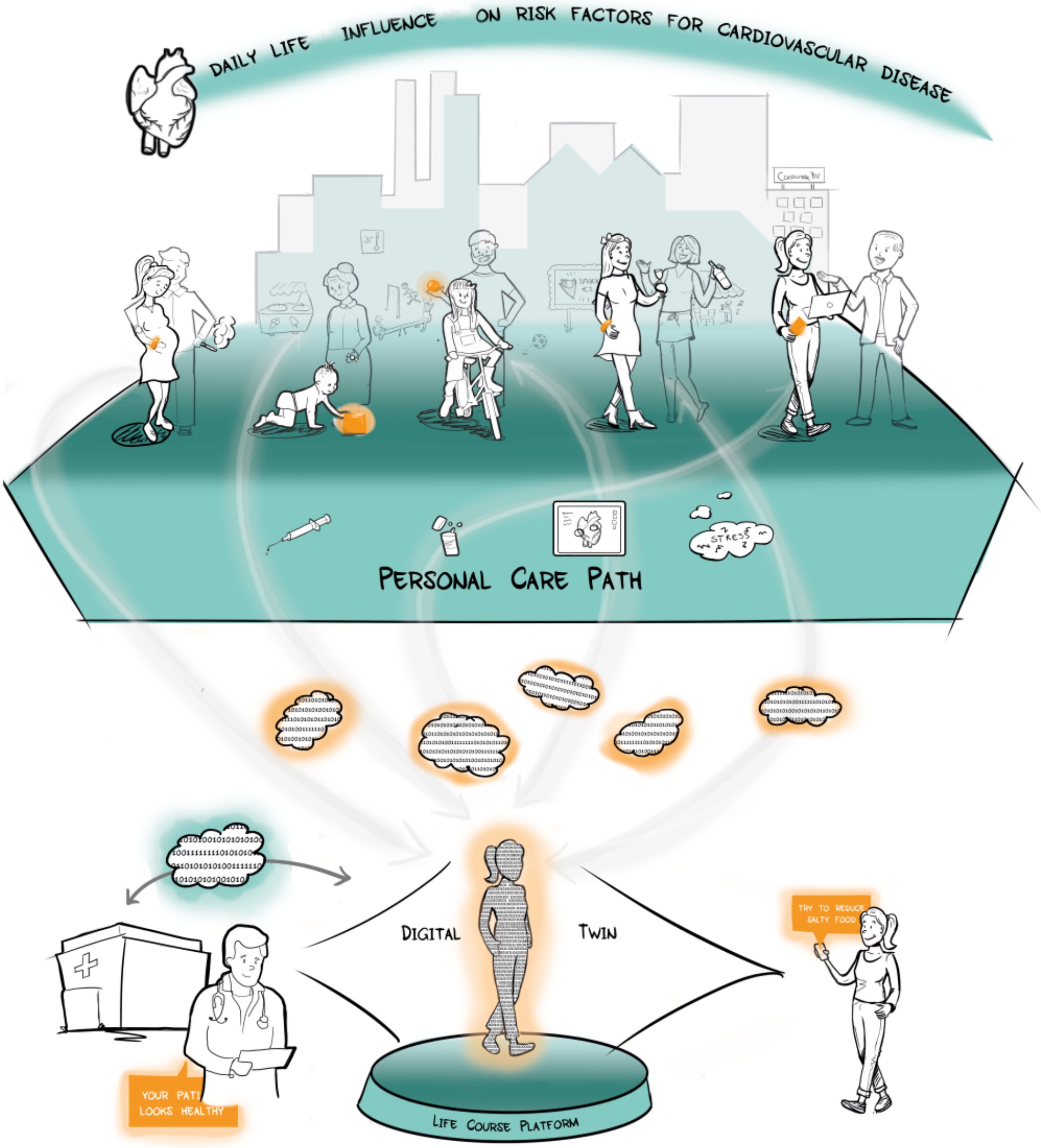
Last, because the medical profession generally deals with interventions, another theme of interest is causal machine learning, where the focus is on creating a model that can predict the change in target output if a predictor were changed (eg, change in diet or blood pressure). The challenging part is inferring causality from observational data, as in medicine we would not only like to associatively predict but intervene in patient health journeys to prevent adverse outcomes. The current gold standard is randomized intervention data, but these data are resource-intensive to collect, and the clinical trial setting does not reflect real life, so the distribution of these data may not fit the general population [32]. However, there are powerful alternatives for causal questioning that may have a future in drug efficacy and safety evaluation on real-world data [33]. Moreover, there are tools being used for evaluating interventions with methods, which can possibly extend into other medical domains [34].

In summary, applying accurate and understandable models on data that fit well to the target individual will be paramount to the success of the DT. Equally important is deconvoluting the causal factors that influence health outcomes, as this knowledge will power interventions based on personalized causal factors. Using a combination of these methods can form the technical analysis part of the DT framework (Figure 2).

## Dynamic Digital Wins

A dynamic DT (Figure 3) is a twin whose relationships—and the data contained within—change over time as the conditions of people’s lives change over time (eg, a young healthy individual who needs preventive care to increase the chance of fertility, a patient treated for a life-threatening disease at the coronary care unit). The dynamic aspect also refers to the changing of the model targets according to the relevant question and value trade-off. As an individual experiences different phases of life, the relevant dynamic DTs are triggered to make a prediction using the up-to-date data. To keep the predictions relevant to the patient, the DT needs to be updated in a continuous manner, preferably with the least amount of burden on the individual. The continuous monitoring of vital signs (for patients in a hospital setting and at home) using wearable devices allows for prompt detection of developing pathologies and early intervention, which may not be possible using standard intermittent vital sign measurements [35]. To have a successful implementation of a dynamic DT, a minimal requirement is to start with a FAIR data set. On this data set, algorithms must be trained and tested, and this output needs to be translated into medical or lifestyle advice. The next step is the validation of this advice to see if this decision-support tool improves patient outcomes. For the full DT, there are multiple requirements: (1) continuous collection of, for example, medical and user experience data, (2) integration of new data with existing data in the databases, and (3) improvement of accessibility of systems that access and store new data [36].

**Figure 3.** A digital twin encompassing different aspects of the life course can be queried to improve medical decision-making to reduce cardiovascular complications. In orange, we show data (and data generators). In green, we display the digital life course platform and real-life course. Together they supply information to each other to support a healthy patient lifestyle by defining a personalized care path.



## Data Management

The first requirement in data management is the collection of high-quality data, which includes similar distribution of patients and individuals as the target individual. As the clinical data is collected from different sources, an important challenge is to synergize, homogenize, integrate, and standardize the heterogeneous data to have a standardized data model. In such a data model, we transform the local data source into the common data model via extract, transform, and load procedures [37]. This data model should live in a system that can sustain FAIR principles. Moreover, patients/researchers and health care providers should have an easy and flexible way to add or remove medical and lifestyle data sources [24]. Data security and privacy are important for DT development in the medicine and health care sectors. Open challenges are cleaning and preparing the data [38] and protecting the privacy of data residing on cloud-based data storage platforms [39]. The rising popularity of storing health data on cloud platforms requires deliberate design in managing data access control and data ownership in combination with schemes such as federated analysis.

## Values

Values are the quality of a digital health service as it is experienced by an individual in relation to their needs [40]. Based on this definition, values are created when users such as patients, clinicians, nurses and others use the DT to address their challenges [41,42]. Representative examples of such challenges were identified by the World Health Organization in 2018 [43] and include efficiency, cost, and access to information.

In our early life course patient journey example, the DT can solve information challenges by collecting high-quality data through wearable devices and then providing easy access and visual summaries of the data to the patient and their treating health care professional. Additionally, the DT can provide personalized predicted probabilities of adverse health outcomes and suggest treatment consisting of lifestyle advice or drug treatment to lower those probabilities. The DT can aid the use challenges of low adherence to treatment by visualizing the effect that treatment may have on future health outcomes so that the patient can better understand and be motivated by the potential benefit of adherence to treatment. The DT addresses efficiency challenges by dynamically updating a patient's predicted probabilities for adverse health outcomes, which enables an early reaction to increased risks. This information may also be used to choose the optimal time between health checks and interventions; in limited care situations, it may be used to prioritize patients with greater needs in triage settings. The cost challenges can be addressed by early risk detection and interventions of the DT, which may prevent the need for costly treatments at later stages of disease progression. To create a valuable DT for users, it may help to include lead users in the development team [24].

## Legal and Ethical Considerations

### *Value Trade-offs*

The DT allows us to optimize for healthy outcomes at the cost of other values (eg, fairness). While incorporating more health-related data might increase a prediction's accuracy, it could also raise concerns about surveillance health care and erode the trust between the health care system and individuals [44]. A DT measuring patient data (such as sleep) and the feedback of health guidance could even affect the parameters it is measuring. For example, commercial consumer wearables are not accurate in estimating sleep [45], and the results may worry consumers and even cause sleep problems. An increased number of measurements may identify patients that are considered atypical. Currently, we lack a clear understanding of how to interpret slight deviations from the normal ranges and means and whether these asymptomatic deviations will lead to future diseases. Balancing the ideal of early diagnosis and reducing overdiagnosis and overtreatment can therefore be challenging [46,47]. The attempt to introduce personalized lifestyle advice to curb lifestyle diseases such as obesity, diabetes, and hypertension also risks promoting an overly individualized view of health management.

Additionally, many have expressed concern about worsening existing health inequity [48,49]. People from lower socioeconomic backgrounds might not have the means to access the digital device required for a DT service, or they might not know how to benefit from the information provided to them due to cognitive constraints such as limited digital or health literacy and external constraints. Instead of empowering users by offering them more health-related information, the DT might burden users with a sense of guilt or anxiety and give rise to the idea that users who do not make the advised change could be accused of being responsible for their adverse health outcomes [50-52].

### *Data Governance and Accountability*

The digitalization of health care gives rise to new legal and accountability issues. For instance, who owns the health data gathered by the DT? Is it owned by the patient, the health care provider, or a third party? Is it morally permissible for a DT service provider to sell user data to a third party? To avoid data being used against the patients' best interest, Schwartz et al [24] suggest 3 principles: (1) patients own their data, which empowers patients to protect the privacy of their data against misuse, (2) patients must provide explicit informed consent for the use of their data, and (3) advocacy efforts should enshrine patient data ownership and access into law. The implementation of these principles poses technical and legislative challenges. Additionally, DTs are susceptible to biases present in the data from which they are developed. Data included in the DT should be reviewed and methods to remedy biases should be considered to avoid the perpetuation of historical biases [24]. De Laat [53] and Nissenbaum [54] describe the obstacles of accountability of machine learning algorithms and how this relates to shared development, human and computer errors, and a culture where it easy to blame the technology. There is also a lack of understanding of how algorithms work [53,54]. Transparency of decision rules and oversight of the decision-making algorithms by governing bodies could solve these outstanding obstacles.

## Scientific Domain

In this paper, the DT was defined in terms of requirements stemming from medical, technological, and ethical domains, but to evaluate the effectiveness of such a DT system, we require scientific convergence to test its merit and limitations (Textbox 1). In translational medical science, we aim to generate knowledge in clinical practice so we can intervene and improve health care processes [55]. To achieve this aim, we need to understand the disease process and uncover the relevant causal pathways that influence health outcomes. DTs can be helpful in combining cross-domain knowledge. As a DT incorporates more than one data set, such as physiological measurements, questionnaires, and lifestyle factors, there are many rich features for each patient or individual ready to be used for pattern recognition. Knowledge about individual contributors of disease progression, in turn, could lead to the identification of subtypes of disease with different disease ontologies and treatments, described as deep phenotyping [13]. The DT platform may allow us to test our predictions more efficiently in an umbrella or real-world trial [16,56]. Predicting patients prospectively and monitoring for outcomes may allow us to test the accuracy of the predictions from the DT. In a second step, randomized intervention based on these predictions can unveil causal factors, and this in turn can help us test our scientific hypotheses using real-world evidence and thereby moving science forward by generating medical knowledge and developing novel technical methodologies.

**Textbox 1.** Benefits of digital twin development in health care.

Health agency and promoting healthy outcomes

- More accurate diagnosis using integrated data
- Improved treatment selection for patients
- Prognosis of patient disease trajectories
- Real-time remote monitoring of health state
- Simulation of treatment and care processes to guide policy

Fair data and validation of research

- Data structures in the digital twin will make data more accessible
- Validation of scientific results provide an implementation framework for decision-making

Algorithm development for medical decision-making, diagnosis, and prevention

- Innovation in algorithms allows us to make more accurate predictions on unseen data
- Causality is inferred from observational data
- Privacy is preserved during data sharing and analysis
- Uncertainty is quantified from different sources and considered during the decision-making processes

## Conclusion

In conclusion, translational science and medical care can be improved by following a DT life course approach: high-dimensional data collection and storage, patient trajectory modeling, outcome predictions, testing, model interpretation, and implementation in clinical practice. By committing multiple nonmedical domains to developing a DT, we aim to improve patient care journeys in a systematic and diligent way.

## Future Research Opportunities

Our vision of a dynamic DT allows for collaboration of many researchers from different domains, where we can align our research applications in order to develop a dynamic DT. Some subjects related to the DT have not been fully investigated, and new opportunities for research have been defined. A part of the technical domain, which is under active investigation, is how we can leverage personal and lifestyle factors to reach more accurate predictions on unseen data. This involves but is not limited to using nonrandom and sparse sampling as a predictor matrix to improve our models and infer causality.

We also need more work on how to identify new classes, such as disease types based on new distance metrics, and how to tune our models to meet patient preferences and different privacy settings. A large body of research in the ethical domain is dedicated to value trade-offs in the DT setting and investigates how we can balance values such as fairness, equality, and health. We are also concerned with data storage and federated data analytics, and we performed analysis on harmonized big data sets. Optimal design of the DT application and user interaction is required to meet the requirements of users while optimizing for values arising from different domains. The science domain should come into play to judiciously evaluate the data, models, application, and ability of the technologies to impact patient life course trajectories and decrease the incidence of disease.

### Acknowledgments

We would like to thank the convergence program, Digital Twin Flagship Erasmus Medical Center and Technical University Delft for leading the technology program [57]. We acknowledge Marina Bos–de Vos for critically reviewing the paper. We thank Emma Schallers for illustrating Figure 3.

### Authors' Contributions

STM, A-HO, AJR-B, K-HK, BB, M Rousian, EB, VV, RH, CA, and RS-T contributed to designing the paper and wrote the first draft of paper. STM, A-HO, AJR-B, and RS-T revised and finalized the draft. All other authors (P-HH, JRvL, SW, PR, DT, M Reinders, and DR) critically reviewed and revised the paper (specifically the parts related to their expertise).

### Conflicts of Interest

None declared.

### References

1. Grieves MW. Product lifecycle management: the new paradigm for enterprises. Int J Prod Devel 2005;2(1/2):71. [doi: 10.1504/ijpd.2005.006669]
2. Roehrs A, da Costa C, Righi R, de Oliveira K. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
3. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917 [FREE Full text] [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
4. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R, HITEC Investigators. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak 2017 Apr 10;17(1):36 [FREE Full text] [doi: 10.1186/s12911-017-0430-8] [Medline: 28395667]
5. Usher-Smith J, Emery J, Hamilton W, Griffin SJ, Walter FM. Risk prediction tools for cancer in primary care. Br J Cancer 2015 Dec 22;113(12):1645-1650 [FREE Full text] [doi: 10.1038/bjc.2015.409] [Medline: 26633558]
6. Dorresteijn J, Visseren F. U-prevent: gepersonaliseerde cardiovasculaire preventie. Huisarts en Wet 2020;63(12):75. [doi: 10.1007/s12445-020-0934-9]
7. Björnsson B, Borrebaeck C, Elander N, Gasslander T, Gawel DR, Gustafsson M, Swedish Digital Twin Consortium. Digital twins to personalize medicine. Genome Med 2019 Dec 31;12(1):4 [FREE Full text] [doi: 10.1186/s13073-019-0701-3] [Medline: 31892363]
8. Alber M, Buganza Tepole A, Cannon WR, Dura-Bernal S, Garikipati K, Karniadakis G, et al. Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. NPJ Digit Med 2019 Nov 25;2(1):115 [FREE Full text] [doi: 10.1038/s41746-019-0193-y] [Medline: 31799423]
9. Corral-Acero J, Margara F, Marciniak M, Rodero C, Loncaric F, Feng Y, et al. The "Digital Twin" to enable the vision of precision cardiology. Eur Heart J 2020 Dec 21;41(48):4556-4564 [FREE Full text] [doi: 10.1093/eurheartj/ehaa159] [Medline: 32128588]
10. Gkouskou K, Vlastos I, Karkalousos P, Chaniotis D, Sanoudou D, Eliopoulos AG. The "Virtual Digital Twins" concept in precision nutrition. Adv Nutr 2020 Nov 16;11(6):1405-1413 [FREE Full text] [doi: 10.1093/advances/nmaa089] [Medline: 32770212]
11. Shamanna P, Saboo B, Damodharan S, Mohammed J, Mohamed M, Poon T, et al. Reducing HbA1c in type 2 diabetes using digital twin technology-enabled precision nutrition: a retrospective analysis. Diabetes Ther 2020 Nov;11(11):2703-2714 [FREE Full text] [doi: 10.1007/s13300-020-00931-w] [Medline: 32975712]
12. Chakshu NK, Carson J, Sazonov I, Nithiarasu P. A semi-active human digital twin model for detecting severity of carotid stenoses from head vibration: a coupled computational mechanics and computer vision method. Int J Numer Method Biomed Eng 2019 May;35(5):e3180 [FREE Full text] [doi: 10.1002/cnm.3180] [Medline: 30648344]
13. Fagherazzi G. Deep digital phenotyping and digital twins for precision health: time to dig deeper. J Med Internet Res 2020 Mar 03;22(3):e16770 [FREE Full text] [doi: 10.2196/16770] [Medline: 32130138]
14. WHO Handbook for Guideline Development. Geneva: World Health Organization URL: https://apps.who.int/iris/bitstream/handle/10665/75146/9789241548441_eng.pdf [accessed 2022-08-26]
15. Cosgrove L, Shaughnessy AF, Shaneyfelt T. When is a guideline not a guideline? The devil is in the details. BMJ Evid Based Med 2018 Feb;23(1):33-36. [doi: 10.1136/ebmed-2017-110845] [Medline: 29367325]
16. Yao X, McCoy RG, Friedman PA, Shah ND, Barry BA, Behnken EM, et al. ECG AI-Guided Screening for Low Ejection Fraction (EAGLE): rationale and design of a pragmatic cluster randomized trial. Am Heart J 2020 Jan;219:31-36. [doi: 10.1016/j.ahj.2019.10.007] [Medline: 31710842]
17. Steegers-Theunissen R, Twigt J, Pestinger V, Sinclair K. The periconceptional period, reproduction and long-term health of offspring: the importance of one-carbon metabolism. Hum Reprod Update 2013;19(6):640-655. [doi: 10.1093/humupd/dmt041] [Medline: 23959022]
18. Alsnes IV, Vatten LJ, Fraser A, Bjørngaard JH, Rich-Edwards J, Romundstad PR, et al. Hypertension in pregnancy and offspring cardiovascular risk in young adulthood. Hypertension 2017 Apr;69(4):591-598. [doi: 10.1161/hypertensionaha.116.08414]
19. Maher G, O'Keeffe G, Kenny L, Kearney P, Dinan T, Khashan A. Hypertensive disorders of pregnancy and risk of neurodevelopmental disorders in the offspring: a systematic review and meta-analysis protocol. BMJ Open 2017 Oct 05;7(10):e018313 [FREE Full text] [doi: 10.1136/bmjopen-2017-018313] [Medline: 28982838]
20. Webster K, Fishburn S, Maresh M, Findlay S, Chappell L, Guideline Committee. Diagnosis and management of hypertension in pregnancy: summary of updated NICE guidance. BMJ 2019 Sep 09;366:l5119. [doi: 10.1136/bmj.l5119] [Medline: 31501137]
21. Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Info Serv Use 2017 Mar 07;37(1):49-56. [doi: 10.3233/isu-170824]
22. Itani W, Kayssi A, Chehab A. Privacy as a service: privacy-aware data storage and processing in cloud computing architectures. 2009 Presented at: 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing; 2009; Chengdu p. 711-716. [doi: 10.1109/dasc.2009.139]

23. Jain A, Duin R, Mao J. Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 2000;22(1):4-37. [doi: 10.1109/34.824819]
24. Schwartz SM, Wildenhaus K, Bucher A, Byrd B. Digital twins and the emerging science of self: implications for digital health experience design and “small” data. Front Comput Sci 2020 Oct 15;2:1-16. [doi: 10.3389/fcomp.2020.00031]
25. Rasheed A, San O, Kvamsdal T. Digital twin: values, challenges and enablers from a modeling perspective. IEEE Access 2020;8:21980-22012. [doi: 10.1109/access.2020.2970143]
26. Richens J, Lee C, Johri S. Author correction: improving the accuracy of medical diagnosis with causal machine learning. Nat Commun 2021 Mar 31;12(1):1-2 [FREE Full text] [doi: 10.1038/s41467-021-21494-9] [Medline: 33790263]
27. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer; 2008.
28. Lee C, Yoon J, Van Der Schaar M. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Trans Biomed Eng 2020 Jan;67(1):122-133. [doi: 10.1109/tbme.2019.2909027]
29. Dempster A, Petitjean F, Webb GI. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. Data Min Knowl Disc 2020 Jul 13;34(5):1454-1495. [doi: 10.1007/s10618-020-00701-z]
30. Reps JM, Rijnbeek P, Cuthbert A, Ryan PB, Pratt N, Schuemie M. An empirical analysis of dealing with patients who are lost to follow-up when developing prognostic models using a cohort design. BMC Med Inform Decis Mak 2021 Feb 06;21(1):43 [FREE Full text] [doi: 10.1186/s12911-021-01408-x] [Medline: 33549087]
31. Wang J, Zhao P, Libol P, Hoi S. Research collection school of information systems. J Mach Learn Res 2014:1 [FREE Full text]
32. Pearl J. Theoretical impediments to machine learning with seven sparks from the causal revolution. 2018. URL: https://ftp.cs.ucla.edu/pub/stat_ser/r475.pdf [accessed 2022-08-26]
33. Kyono T, Zhang Y, Van Der Schaar M. CASTLE: regularization via auxiliary causal graph discovery. URL: http://proceedings.neurips.cc/paper/2020/file/1068bceb19323fe72b2b344ccf85c254-Paper.pdf [accessed 2022-08-26]
34. Berrevoets J, Jordon J, Gimson A, Van Der Schaar M. OrganITE: optimal transplant donor organ offering using an individual treatment effect. URL: https://www.vanderschaar-lab.com/papers/NeurIPS2020_OrganITE.pdf [accessed 2022-08-26]
35. Leenen J, Leerentveld C, van Dijk J, van Westreenen H, Schoonhoven L, Patijn G. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. J Med Internet Res 2020 Jun 17;22(6):e18636 [FREE Full text] [doi: 10.2196/18636] [Medline: 32469323]
36. Rivera L, Villegas N, Jiménez M, Tamura G, Angara P, Müller H. Towards continuous monitoring in personalized healthcare through digital twins. CASCON '19: Proc 29th Ann Int Conf Comput Sci Softw Engin 2019:329-335.
37. Bogatu A. Cost-effective data wrangling in data lakes. 2020. URL: https://www.research.manchester.ac.uk/portal/files/162160622/FULL_TEXT.PDF [accessed 2022-08-26]
38. Fuller J. Impact of medical device integration on real-time visibility of electronic medical records. URL: https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=11392&context=dissertations [accessed 2022-08-26]
39. Barricelli B, Casiraghi E, Fogli D. A survey on digital twin: definitions, characteristics, applications, and design implications. 2019. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8901113 [accessed 2022-08-26]
40. Lepak D, Smith K, Taylor M. Value creation and value capture: a multilevel perspective. Acad Manag Rev 2007 Jan;32(1):180-194. [doi: 10.5465/amr.2007.23464011]
41. Vargo S, Lusch R. Evolving to a new dominant logic for marketing. Serv Log Mark Dialog Debate Dir 2014;68:28. [doi: 10.4324/9781315699035-9]
42. Lusch R, Nambisan S. Service innovation: a service-dominant logic perspective. MIS Q 2015 Jan 1;39(1):155-175 [FREE Full text] [doi: 10.25300/misq/2015/39.1.07]
43. WHO classification of digital health interventions. 2018. URL: http://www.who.int/reproductivehealth/publications/mhealth/classification-digital-health-interventions/en/ [accessed 2022-08-26]
44. Terrasse M, Sisti D. Policing compliance: digital medicine and criminal justice-involved persons. Am J Bioeth 2018 Sep;18(9):57-58 [FREE Full text] [doi: 10.1080/15265161.2018.1498937] [Medline: 30235101]
45. Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. Sleep Med Rev 2020 Feb;49:101227. [doi: 10.1016/j.smrv.2019.101227]
46. Slow Medicine: the Apple Watch and overdiagnosis. URL: https://www.medpagetoday.com/opinion/slowmedicine/75946 [accessed 2021-07-21]
47. Duffy D. Problems, challenges and promises: perspectives on precision medicine. Brief Bioinform 2016 May;17(3):494-504. [doi: 10.1093/bib/bbv060] [Medline: 26249224]
48. Ibrahim H, Liu X, Zariffa N, Morris A, Denniston A. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021 Apr;3(4):e260-e265 [FREE Full text] [doi: 10.1016/S2589-7500(20)30317-4] [Medline: 33678589]
49. Cornejo Müller A, Wachtler B, Lampert T. [Digital divide-social inequalities in the utilisation of digital healthcare]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Feb;63(2):185-191 [FREE Full text] [doi: 10.1007/s00103-019-03081-y] [Medline: 31915863]

50. Marantz P. Blaming the victim: the negative consequence of preventive medicine. Am J Public Health 1990 Oct;80(10):1186-1187. [doi: 10.2105/ajph.80.10.1186] [Medline: 2400027]
51. Owens J, Cribb A. "My Fitbit thinks i can do better!" Do health promoting wearable technologies support personal autonomy? Philos Technol 2017 Jun 6;32(1):23-38. [doi: 10.1007/s13347-017-0266-2]
52. Prainsack B. Personalized Medicine: Empowered Patients in the 21st Century?. New York: NYU Press; 2017.
53. de Laat P. Algorithmic decision-making based on machine learning from big data: can transparency restore accountability? Philos Technol 2018;31(4):525-541 [FREE Full text] [doi: 10.1007/s13347-017-0293-z] [Medline: 30873341]
54. Nissenbaum H. Accountability in a computerized society. Sci Eng Ethics 1996 Mar;2(1):25-42. [doi: 10.1007/bf02639315]
55. Drucker D. Never waste a good crisis: confronting reproducibility in translational research. Cell Metab 2016 Sep 13;24(3):348-360 [FREE Full text] [doi: 10.1016/j.cmet.2016.08.006] [Medline: 27626191]
56. Park J, Siden E, Zoratti M, Dron L, Harari O, Singer J, et al. Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. Trials 2019 Sep 18;20(1):572 [FREE Full text] [doi: 10.1186/s13063-019-3664-1] [Medline: 31533793]
57. Digital Twin. URL: https://convergence.nl/health-technology/digital-twin/ [accessed 2022-08-31]

## Abbreviations

**DT:** digital twin
**FAIR:** findable, accessible, interoperable, and reusable

*Edited by A Mavragani; submitted 14.12.21; peer-reviewed by H Luo, J Heo; comments to author 18.01.22; revised version received 31.05.22; accepted 09.06.22; published 14.09.22.*

*Please cite as:*
*Mulder ST, Omidvari AH, Rueten-Budde AJ, Huang PH, Kim KH, Bais B, Rousian M, Hai R, Akgun C, van Lennep JR, Willemsen S, Rijnbeek PR, Tax DMJ, Reinders M, Boersma E, Rizopoulos D, Visch V, Steegers-Theunissen R*
*Dynamic Digital Twin: Diagnosis, Treatment, Prediction, and Prevention of Disease During the Life Course*
*J Med Internet Res 2022;24(9):e35675*
*URL: https://www.jmir.org/2022/9/e35675*
*doi:10.2196/35675*
*PMID:36103220*

Viewpoint

# On the Current Connection and Relation Between Health Informatics and Social Informatics

Zdenek Smutny[1*], PhD; Vasja Vehovar[2*], PhD

[1]Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic

[2]Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia

[*]all authors contributed equally

**Corresponding Author:**
Zdenek Smutny, PhD
Faculty of Informatics and Statistics
Prague University of Economics and Business
W. Churchill Sq. 1938/4
Prague, 13067
Czech Republic
Phone: 420 224095473
Email: zdenek.smutny@vse.cz

## Abstract

Scholars from the health and medical sciences have recently proposed the term social informatics (SI) as a new scientific subfield of health informatics (HI). However, SI is not a new academic concept; in fact, it has been continuously used in the social sciences and informatics since the 1970s. Although the dominant understanding of SI was established in the 1990s in the United States, a rich international perspective on SI has existed since the 1970s in other regions of the world. When that perspective is considered, the fields of understanding can be structured into 7 SI schools of thought. Against that conceptual background, this paper contributes to the discussion on the relationship between SI and HI, outlining possible perspectives of SI that are associated with health, medical, and clinical aspects. This paper argues against the multiplication and inconsistent appearance of the term SI when newly used in health and medical sciences. A more explicit name for the area that uses health and social data to advance individual and population health might be helpful to overcome this issue; giving an identity to this new field would help it to be understood more precisely and bring greater separation. This labeling could be fruitful for further segmentation of HI, which is rapidly expanding.

***(J Med Internet Res 2022;24(9):e40547)*** doi:10.2196/40547

**KEYWORDS**

biomedical informatics; conceptual view; clinical informatics; international perspective; medical informatics

## Introduction

Social informatics (SI) refers to research activities related to the interaction between information and communication technology (ICT) and modern society. SI has been continuously defined and used in various regions and contexts since the early 1970s [1]. A broad international discourse has already been overviewed [2], where 7 (US, German, Russian, Norwegian, Japanese, Slovenian, and UK) regional SI schools of thought were identified. The US school, founded by the pioneering work of Rob Kling (1944-2003) [3], is globally dominant, although a detailed overview of the SI literature showed that it is steadily declining [4].

In addition to these contributions, Pantell et al [5] recently proposed a new and specific understanding of the term SI as a subfield of health informatics (HI), "focused on the application of information technologies to capture and apply social data in conjunction with health data to advance individual and population health" [5]. Shachak [6] challenged whether SI is an appropriately chosen label here, given its establishment in the US school of SI dating back to the mid-1990s [3,7,8]. This stimulated Pantell et al [9] to provide further arguments on why SI is a suitable proposed name. We would like to contribute to this discussion from a broader perspective than that specific to the US school of SI [6] by introducing the international and multidisciplinary aspects of SI. This paper is intended for the health community. It, therefore, presents the detailed conceptual basis of SI, which is not as well-known in this community as the conceptual basis of HI [10].

## Development and Perspectives of Social Informatics

When considering the interaction between ICT and society, SI addresses the social aspects of ICT as well as ICT aspects of social issues, including ICT applications in social sciences [2]. However, there is no unified general understanding of this interaction, and numerous well-established research fields overlap and compete here: computational social science, information systems, information society, science and technology studies, ICT and society, internet studies, information sociology, computer-mediated communication, human-computer interaction, and many others. There are also 2 main perspectives in SI, which differ substantially. The first is a regional perspective, which is linked to the theoretical and conceptual backgrounds of the 7 SI schools [11,12]. The second is a broad international perspective, which encompasses the relevant thematic areas in which the self-declared term SI can appear with reference to academic and scientific discourse [2,13].

The research that declares itself SI covers only a portion of this interaction, and it is spread across various thematic areas. These thematic areas where SI can appear are broad but can be classified into 3 streams according to their scientific origin [2,14]:

1. *Social science*, where the social aspects of ICT can be observed at a personal, organizational, or societal level.
2. *Informatics* (including computer science, information science, and information systems), where ICT is applied in the social area, including the provision of public or business services. This stream also includes computer modeling; artifact or solution design; and the structuring, conceptualization, and processing of information (eg, information architecture, information visualization, or information design).
3. *Social science methodology*, where ICT can be involved in various stages of the social science research process: data collection, processing, and analysis. The ICT applications are understood here as tools for social science research.

The above scope of SI is thus much broader than the domain-specific application of informatics (stream 2) in social sciences (eg, business informatics, political science informatics, music informatics, and legal informatics) and broader than the sole social science perspective on SI (stream 1). The latter is actually one of the main focuses of the US school of SI, which concerns “the interdisciplinary study of the design, uses and consequences of information technologies that takes into account their interaction with institutional and cultural contexts” [3]. The US school of SI addresses the social aspects of ICT, as well as ICT aspects of social issues: “SI refers to the study of social aspects of computerization, including the role of information technology in social and organizational change, the use of information technologies in social contexts, and the way that the social organization of information technologies is influenced by social forces and social practices” [15]. Contextual grasping is typical for SI (not only in the United States) and can include cultural, institutional, social, ethical, legal, or other issues. SI is defined thematically and not methodologically [16], so SI is open to the use of different research methods. Recent review papers [2,4,13] have confirmed that the prevailing scientific contributions in SI, referring to any of the SI schools, remain conceptually strongly rooted in the social sciences (stream 1), despite the declining dominance of the US school.

Since the 1970s [1,14,17], numerous scholars have thus used the term SI for their research, addressing the interaction between ICT and society. Given the above-described broadness of the potential SI scope, it is not surprising that considerable differences exist in the academic communities between various regions or schools of SI. However, the introduction of a different understanding of scientific terms is not a new phenomenon. A typical example is already the term *informatics*, dating back to the 1950s and especially the 1960s. Currently, the most widespread understanding of informatics is the conception originating in France [18-21], which recognizes informatics as a broad computer-oriented discipline that covers information and computational processes, including the logical construction of computer systems. Due to historical circumstances, a different understanding of informatics evolved in the United States [22-24] because of the strong existence of previously established names of computer-oriented disciplines, such as computer science, computer engineering, information systems, or library and information science. The US understanding of informatics is thus referred to more narrowly as a specific computer application–oriented domain with a user-centered perspective [25] and primary rooted in (library and) information science. The main emphasis is on the nonnumerical use of computers, and the central point of interest is information and its processing [26].

Within this context, the history of the term *medical informatics* in the United States is also informative for the understanding of informatics in the United States. Thanks to an action of the International Federation for Information Processing (Technical Committee 4 on medical informatics), the term was promoted and later adopted in the United States in the 1970s [27-29]. The term medical informatics thus existed independently in the United States for many years without referring to the general notion of informatics or the existence of other domain-applied informatics—that is, before the stand-alone term “informatics” was generally accepted in the United States.

It should be added that according to Hersh [26], in the United States, the following terms are subsumed under the broad umbrella *biomedical and health informatics* (sometimes only components are used, such as *biomedical informatics* or *HI*): *bioinformatics* (cellular and molecular level), *medical* and *clinical informatics* (personal level), and *public health informatics* (population level). Relevant SI research may appear from the international view of SI and the perspective of Pantell et al [5] in areas of medical (and clinical) informatics and population health informatics.

Nevertheless, as presented in Table 1, all existing schools of SI, including the US one, agree that SI relates to the general interaction (or combination) of society with ICTs [2], the computerization of human activities [8], or the social aspects of computerization [3], where the term “social” relates to social sciences in general.

**Table 1.** Main thematic focus of 7 SI schools. Based on and modified from Smutny and Vehovar [2].

| SI[a] school | Origin in existing disciplines | Typical research focus | Basic sources |
|---|---|---|---|
| German school | Interdisciplinary informatics with sociology, economics, psychology, and social work and services | Interactions of software with individuals, organizations, and society; the use of information and communication technology in social work | Sozioinformatik: Wulf et al, 2018 [30]; and Sozialinformatik: Kreidenweis [31], 2012 |
| Japanese school | Informatics, media, and communication studies | Study of social information, communication, and information processes in an information society | Kurosu [32], 2010 |
| Norwegian school | Sociology, psychology, and sociotechnical research | Study of the design, deployment, and evaluation of information and knowledge systems | Maliţa [33], 2006; and Bråten [34], 1983 |
| Russian school | Initially, library science and communication studies; currently, informatics, sustainable development, and the philosophy of information, education, and knowledge management | Interaction between society and ICT[b], along with sustainable development, with strong educational and philosophical overlaps | Chugunov [35], 2012; Kolin [17], 2021; and Melnikova and Romanovskaya [36], 2021 |
| Slovenian school | Sociology, statistics, and informatics | Use of ICT to study the interactions between ICT and society | Petric and Atanasova [37], 2013; and Vehovar [38], 2006 |
| UK school | Sociology, information science, and science and technology studies | Transdisciplinary study of sociotechnical interactions | Davenport [39], 2008 |
| US school | Information science, information systems, sociology, and computer science | Study of the design, uses, and consequences of information technologies in institutional and cultural contexts | Kling [3], 2007; and Fichman et al [7], 2015 |

[a] SI: social informatics.

[b] ICT: information and communication technology.

However, we should also note that various language-specific exceptions exist here. For example, the German understanding of SI, according to the German term “sozialinformatik” [40], focuses on the application of ICTs in social work and services, as the term “sozial” in German has a strong connotation with social work. Nevertheless, this is, in a large part, a specific aspect of some Germanic languages. Therefore, in Scandinavia [41,42] and Germany, SI sometimes focuses on the social, educational, and health sectors [40,43], although this is also not a general or prevailing understanding of SI in these countries. We may add that the notion of SI, in general, frequently suffers from various translation issues, arising from the ambiguity and vagueness of the notions of “informatics” and “social,” which are both already problematic (ie, they have multiple meanings) in English. Therefore, SI is sometimes (back)-translated to English as socioinformatics (Germany), socio-information studies (Japan), or social information science (China) [2,14,44].

In addition, the term SI is increasingly used as a self-declared general label in the second stream of SI, particularly in connection to so-called computational social science, where ICTs are applied in various social science domains [4]. The related occurrence of the SI term in databases and searches (Web of Science [WoS], Scopus, and Google Search) and its usage to label various activities associated with the interaction between the ICT and society are growing exponentially, much faster than scholarly literature on SI referring to any of the SI schools [4]. This expansion of the usage of the term SI is outside of any SI school; it is actually very surprising, and in a way, also problematic, because it stems from the superficial usage of the term SI and has no conceptual grounding and thus is hard to explain.

We can thus summarize that SI is a very specific attempt to encompass the research related to the interaction between ICT and society. For various reasons, SI did not gain general popularity, in part because of the numerous alternative scientific fields that also address this interaction, and in part due to language issues, but also due to its internal diversity across 7 schools of SI, which span well beyond the US school. The recent expansion of the usage of the term SI outside scholarly production and outside any conceptual grounding makes this scenario even more problematic. Still, SI is a well-articulated scientific field, which addresses the general (ie, not field-specific) scope of the interaction between ICT and society.

## Health Informatics and Social Informatics

The alternative definition of SI proposed by Pantell et al [5] focuses only on a specific health-oriented community—for instance, in their argumentation, they used only the MEDLINE database for the search [9]—thus being defined as a subfield of HI. Nevertheless, in principle, the situation can be conceptually reversed from the broadest international perspective of SI, where HI can be subsumed under SI [2] as a domain-specific application of ICTs in society (stream 2).

To present the broad international and multidisciplinary view of SI and its connection with HI, we referred to the world’s most recognized citation databases, WoS and Scopus. Notably, there is a limitation in that they only focus on English terms, and documents in regional languages are omitted (eg, eLibrary.ru [Russia] and J-Stage [Japan]).

In the following comparison, we focused on documents located in the WoS and Scopus citation databases, where a search term appears in the title, abstract, or keywords. The search used the following terms: (TITLE-ABS-KEY (“social informatics”) AND NOT TITLE-ABS-KEY (“Institute for Computer Sciences, Social Informatics”)); (TITLE-ABS-KEY (“health informatics”)); “social informatics” (Topic) not “Institute for Computer Sciences, Social Informatics” (Topic); and “health informatics” (Topic). We removed articles that only mentioned *Lecture Notes of the Institute Computer Sciences, Social Informatics, and Telecommunications Engineering* in the abstract, without a connection to SI. The results for the last decade are presented in Table 2.

**Table 2.** Comparison of search terms in the Web of Science and Scopus citation databases during 2012-2021 (source: Web of Science and Scopus search on February 8, 2022).

| Year | Web of Science (documents), n | | Scopus (documents), n | |
|---|---|---|---|---|
| | Health informatics | Social informatics | Health informatics | Social informatics |
| 2012 | 169 | 7 | 262 | 12 |
| 2013 | 203 | 8 | 286 | 19 |
| 2014 | 171 | 8 | 283 | 24 |
| 2015 | 233 | 3 | 346 | 13 |
| 2016 | 228 | 7 | 335 | 11 |
| 2017 | 270 | 3 | 422 | 7 |
| 2018 | 256 | 12 | 382 | 14 |
| 2019 | 354 | 19 | 508 | 26 |
| 2020 | 385 | 6 | 545 | 15 |
| 2021 | 410 | 10 | 612 | 12 |
| Overall | 2679 | 83 | 3981 | 153 |

Here, the number of documents in the citation databases (WoS and Scopus) is given in parentheses. The number of documents in the last decade comprises 83 (WoS) and 153 (Scopus) SI documents and 2679 (WoS) and 3981 (Scopus) HI documents. The situation in terms of the total number of documents found (not only in the last decade) is similar, confirming that the SI output (WoS: n=210; Scopus: n=325) is very small compared to HI (WoS: n=3870; Scopus: n=6387). If we extend the search to nearby HI areas, such as medical informatics and clinical informatics, the number of documents increases (WoS: n=8894; Scopus: n=38,288).

From the total number of documents (WoS: n=210; Scopus: n=325) with SI in the title, abstract, or keywords, 16 (WoS) and 32 (Scopus) of these documents also refer to the following terms: health, medical, or clinical informatics. This finding implies that from an international perspective, SI research (stream 2), in part, overlaps thematically with HI and related fields. More evidence that SI research appears in the health domain can be found in Cech [45] and Hoeffner [46]. For example, Cech [45] focused on SI research and educational institutions and identified 5 main themes of SI from an international view, and 1 of them was eHealth. We should also add that throughout the review of all scientific publications that explicitly refer to SI [2,4,14], an entry that explicitly refers to SI as a subfield of HI was not found.

Similarly, the number of documents explicitly referring to SI in the MEDLINE database is almost negligible (6 papers), as stated by Pantell et al [9], in addition to 58 papers where SI is mentioned only through an affiliation. Notably, these 58 affiliations are likely referring to the standard SI understanding (and not to that in the context of the proposed HI subfield).

We can conclude that currently, roughly around one-tenth of SI scholarly publications refer to the areas related to HI, whereas, for now, the scholarly literature on HI rarely mentions the notion of SI (although it somehow often refers to SI through an affiliation, which is related to the standard SI understanding; ie, the general interaction between ICT and society, which is not narrowed to HI).

## Discussion

Various conceptually focused scholarly pieces of literature exist on SI [2,4,7,13]. Correspondingly, 7 SI schools of thought can be identified, with the US school (founded by Rob Kling and his colleagues in the 1990s) being the most globally visible, although its impact is steadily declining. However, the US school of SI, which is in large rooted in social sciences (stream 1)—and can be, with some simplification, denoted as a study of the social aspects of computerization—is constantly evolving [7,47], and scholars are adapting to current research trends relevant to other areas as well (eg, computational social sciences and sociotechnical systems research).

From an international and interdisciplinary perspective, one can argue that a certain commonly established concept, such as SI, cannot be replicated in a very specific academic community—especially in socioconstructivist scientific disciplines, such as informatics—without soundly elaborating on the relations to other scientific fields. The multidisciplinary nature of modern science suggests that at least minimal consistency must be preserved across disciplines. The existing definitions of informatics and SI sometimes vary, which is already confusing. However, they still address the interaction between ICT and society in general, although they stem from different historical, ideological, and methodological circumstances or language-specific aspects. Therefore, to avoid further (and radical) increases in inconsistencies—which would be, in this specific case, produced with full awareness and intention—we suggest that for the new SI as a subfield of HI [5], at least an elementary conceptualization must be made with reference to the existing international perspective associated with the established notion of SI. This conceptualization would increase the precision of the term and decrease confusion, particularly if the new term itself is already reflected in the corresponding hierarchy.

Consequently, due to the complexity and specifics of this newly proposed field [5,9], perhaps 3 or more words would be needed. Within this context, besides the terms *social determinants of health informatics* and *clinical social informatics*, which were already considered but rejected because of awkwardness [5], the potential combinations that could perhaps be considered include *health social informatics*, *social informatics in health*, *social health informatics*, *medical social informatics*, or *social informatics in medicine*. In particular, *health social informatics* astutely refers to the health and social perspectives' connection.

We may add that the usage of a 3-keyword description (eg, *health social informatics*) is not only an unambiguous solution to denote a complex and specific field but also reflect similar practice in many other areas where nuances and precision are needed. An excellent example is the established notion of *public health informatics* [48], as well as the more newly proposed neighboring notion of *population health informatics* suggested by Kharrazi et al [49]. The latter is more general than public health informatics and overlaps with other informatics fields, such as public health informatics, clinical informatics, and consumer health informatics.

When addressing the fact that the term SI is already in use with different meanings, Pantell et al [9] justified their proposal by arguing that some terms often have more than 1 meaning anyway, stating the example of *medical sociology*, which covers *sociology of medicine* and *sociology in medicine*. However, this example [9] cannot be used in this situation, where an already established and well-elaborated term (SI) is replicated in another context and given an entirely different meaning. The stated example of medical sociology would be relevant only if the authors, for example, discussed whether SI encompasses *sociology of informatics* and *sociology in informatics*, or *social sciences of informatics* and *social sciences in informatics*, but not for the case where specific domain-applied informatics (from stream 2) declares its SI component as a stand-alone notion of SI. The existence of multiple meanings of SI in 1 region, such as the United States, will not only confuse scholars at the regional level (followers of Kling [3] and followers of Pantell et al [9]) but this confusion will be transferred to the international level as English is the primary language in scientific publishing.

The case of SI is particularly critical because almost all scientific disciplines have certain informatics components and certain social science components. Using the general notion of SI to denote the narrow contexts of components within a certain discipline (ie, HI) would clearly lead to multiplication and inconsistent appearance of the term SI across disciplines; it would also increase the assumption that scientific disciplines are entirely independent and isolated. Most of all, this notion would neglect the existing, established, and generally accepted broader notion of SI. With such an approach, for example, *music informatics*, one of many domain-specific applications of informatics (stream 2), may also directly use the term SI as its subfield when "information technologies are used to capture and apply social data in conjunction to music data."

Beyond this notion, Pantell et al [9] justified their proposal by referring to how, currently, the term SI is rarely used in the PubMed database. However, there are 67 publications already using this notion, mostly as an affiliation with the general notion of SI, which is not a negligible number. With the proposed usage of SI as a subfield of HI, the scholars addressing social determinants of HI will increasingly encounter SI publications from social sciences and informatics, which are related to the existing (general) notion of SI (particularly streams 1 and 2) and vice versa. This usage will create unnecessary confusion and inconsistency within and across disciplines. Similarly, the SI affiliations will lose clarity, because the new usage of the term SI will mostly contradict the SI affiliations, which typically refer to the established meaning. It will also become increasingly ambiguous as to which meaning of SI is to be understood by an SI affiliation.

We should also warn that the US school of SI, which Pantell et al [5,9] and Shachak [6] exclusively refer to, is not the only research stream where SI appears. It is true that for various reasons, it still dominates in a number of scholarly publications, but its role is steadily declining [4]. In any case, in scholarly literature, the stand-alone notion of SI already has various established regional meanings, which basically all address the general interaction between ICT and society, not just a particular domain-specific application of informatics. In addition, we should also beware of the exponential usage of the term SI outside conceptually grounded scholarly literature, which may further increase confusion if another usage of SI appears.

A more explicit and descriptive name—we already identified some potential 3-word options above—could avoid all these problems, bringing much more separation and precision, without any awkwardness or losses when denoting a new discipline that focuses "on the application of information technologies to capture and apply social data in conjunction with health data to advance individual and population health" [5]. Such labeling could be highly fruitful also for further segmentation of HI, which is rapidly expanding and needs more structuring in subdisciplines, together with the corresponding labeling.

## Acknowledgments

The authors received public financial support from the Slovenian Research Agency (P5-0399). This work was also supported by the Faculty of Informatics and Statistics of Prague University of Economics and Business (IP400040).

## Conflicts of Interest

None declared.

## References

1. Hawgood J. Informatics is a social science. Personnel Review 1973 Feb 01;2(2):82-87. [doi: 10.1108/eb055233]
2. Smutny Z, Vehovar V. Social informatics research: schools of thought, methodological basis, and thematic conceptualization. J Assoc Inf Sci Technol 2020 May;71(5):529-539. [doi: 10.1002/asi.24280]
3. Kling R. What is social informatics and why does it matter? The Information Society 2007 Jun 21;23(4):205-220. [doi: 10.1080/01972240701441556]
4. Vehovar V, Smutny Z, Bartol J. Evolution of social informatics: publications, research, and educational activities. The Information Society 2022 Jul 07:1-27. [doi: 10.1080/01972243.2022.2092570]
5. Pantell MS, Adler-Milstein J, Wang MD, Prather AA, Adler NE, Gottlieb LM. A call for social informatics. J Am Med Inform Assoc 2020 Nov 01;27(11):1798-1801 [FREE Full text] [doi: 10.1093/jamia/ocaa175] [Medline: 33202020]
6. Shachak A. Social informatics is a poor choice of term: a response to Pantell et al. J Am Med Inform Assoc 2021 Jun 12;28(6):1356-1357 [FREE Full text] [doi: 10.1093/jamia/ocab021] [Medline: 33659981]
7. Fichman P, Sanfilippo MR, Rosenbaum H. Social Informatics Evolving. Cham: Springer; 2015.
8. Sawyer S, Eschenfelder KR. Social informatics: perspectives, examples, and trends. Ann Rev Info Sci Tech 2005 Feb 01;36(1):427-465 [FREE Full text] [doi: 10.1002/aris.1440360111]
9. Pantell MS, Adler-Milstein J, Wang MD, Prather AA, Adler NE, Gottlieb LM. A reply to Shachak. J Am Med Inform Assoc 2021 Jun 12;28(6):1358-1359 [FREE Full text] [doi: 10.1093/jamia/ocab022] [Medline: 33659993]
10. Mantas J. Biomedical and health informatics education - the IMIA years. Yearb Med Inform 2016 Aug 02;Suppl 1:S92-S102 [FREE Full text] [doi: 10.15265/IY-2016-032] [Medline: 27488405]
11. Rosenbaum H. Social informatics as a scientific and intellectual movement. In: Fichman P, Rosenbaum H, editors. Social Informatics: Past, Present and Future. Cambridge, United Kingdom: Cambridge Scholars Publishing; 2014:2-28.
12. Smutny Z. Social informatics as a concept: widening the discourse. J Inf Sci 2016 Jul 10;42(5):681-710. [doi: 10.1177/0165551515608731]
13. Vehovar V, Smutny Z, Robbin AR. What is social informatics from an international perspective? Acta Inform Pragensia 2021 Dec 31;10(3):207-210. [doi: 10.18267/j.aip.173]
14. Smutny Z. Diskurz sociální informatiky : metodologická východiska, myšlenkové školy a tematická konceptualizace. Discourse of Social Informatics: Methodological Backgrounds, Schools of Thought and Thematic Conceptualization. Prague, Czechia: Professional Publishing; 2020.
15. Rob Kling Center for Social Informatics. Indiana University. URL: https://rkcsi.luddy.indiana.edu/ [accessed 2022-05-20]
16. Nakata K. A cognitive perspective on social informatics. International Journal of Social and Humanistic Computing 2008 Sep 29;1(1):28-35. [doi: 10.1504/ijshc.2008.020478]
17. Kolin KK. Social informatics: 30 years of development of Russian scientific school. Acta Inform Pragensia 2021 Dec 31;10(3):289-300. [doi: 10.18267/j.aip.150]
18. Mounier-Kuhn P, Pégny M. AFCAL and the emergence of computer science in France: 1957-1967. 2016 Jun 14 Presented at: CiE 2016: Pursuit of the Universal; June 27 to July 1, 2016; Paris, France p. 170-181. [doi: 10.1007/978-3-319-40189-8_18]
19. Garric D. L'Informatique: Révolution totale. Paris, France: Publications premières; 1969.
20. Coy W. Defining discipline. In: Foundations of Computer Science. Berlin: Springer; May 16, 1997:21-35.
21. L'Informatique entre a L'Académie française. Le Monde. 1967 Apr 22. URL: http://www.lemonde.fr/archives/article/1967/04/22/l-informatique-entre-a-l-academie-francaise_2625363_1819218.html [accessed 2022-05-20]
22. Bauer WF. Computer recollections: events, humor, and happenings. IEEE Annals Hist Comput 2007 Jan;29(1):85-89. [doi: 10.1109/mahc.2007.2]
23. Friedman CP. What informatics is and isn't. J Am Med Inform Assoc 2013 Mar;20(2):224-226 [FREE Full text] [doi: 10.1136/amiajnl-2012-001206] [Medline: 23059730]
24. Collen MF, Ball MJ. The History of Medical Informatics in the United States. London, UK: Springer; 2015.
25. Detail for CIP Code 11.0104. National Center for Education Statistics. URL: https://nces.ed.gov/ipeds/cipcode/cipdetail.aspx?y=56&cipid=91492 [accessed 2022-05-20]
26. Hersh W. A stimulus to define informatics and health information technology. BMC Med Inform Decis Mak 2009 May 15;9(1):24 [FREE Full text] [doi: 10.1186/1472-6947-9-24] [Medline: 19445665]
27. Roger France F. About the beginnings of medical informatics in Europe. Acta Inform Med 2014 Feb;22(1):11-15 [FREE Full text] [doi: 10.5455/aim.2014.22.11-15] [Medline: 24648614]

28. Masic I. The history of medical informatics development - an overview. International Journal on Biomedicine and Healthcare 2020;8(1):37 [FREE Full text] [doi: 10.5455/ijbh.2020.8.37-52]
29. Degoulet P, Fieschi M, Goldberg M, Salamon R. François Grémy, a humanist and information sciences pioneer. Yearb Med Inform 2014 Aug 15;9:3-5 [FREE Full text] [Medline: 25123715]
30. Wulf V, Pipek V, Randall D, Rohde M, Schmidt K, Stevens G. Socio-informatics. Oxford, United Kingdom: Oxford University Press; May 08, 2018.
31. Kreidenweis H. Lehrbuch Sozialinformatik. Baden, Germany: UTB; 2012.
32. Kurosu T. In search of a paradigm of socio-informatics: on socio-informatics and social informatics. Journal of Socio-Informatics 2010;3(1):69-81. [doi: 10.14836/jsi.3.1_69]
33. Maliţa L. Interviu cu Ingar Roggen. Revista de Informatica Sociala 2006;3:4-11.
34. Bråten S. Dialogens vilkår i datasamfunnet: essays om modellmonopol og meningshorisont i organisasjons- og informasjonssammenheng. Oslo, Norway: Universitetsforlaget; 1983.
35. Chugunov AV. Социальная информатика: Учебное пособие. Saint Petersburg, Russia: НИУ ИТМО; 2012.
36. Melnikova NI, Romanovskaya OA. The Russian concept of social informatics in light of information technology innovation: a systematic review. Acta Inform Pragensia 2021 Dec 31;10(3):301-332. [doi: 10.18267/j.aip.172]
37. Petrič G, Atanasova S. Družboslovna informatika: Razvojne konvergence in raziskovalni dosežki. Teorija in praksa 2013;50(2):375-375 [FREE Full text]
38. Vehovar V. Social informatics: an emerging discipline? 2006 Presented at: HCC 2006: Social Informatics: An Information Society for all? In Remembrance of Rob Kling; September 21-23, 2006; Maribor, Slovenia p. 73-85. [doi: 10.1007/978-0-387-37876-3_6]
39. Davenport E. Social informatics and sociotechnical research — a view from the UK. Journal of Information Science 2008 Jun 13;34(4):519-530. [doi: 10.1177/0165551508091011]
40. Janatzek U. Sozialinformatik - eine wissenschaftstheoretische Verortung. Standpunkt: sozial 2013 Mar;24:36-45 [FREE Full text]
41. Smaradottir BF, Gerdes MW. Teaching methods in a health and social informatics Master program. New York: IEEE; 2019 Presented at: 2019 International Conference on Computational Science and Computational Intelligence; December 5-7, 2019; Las Vegas, NV p. 888-891. [doi: 10.1109/csci49370.2019.00169]
42. Häyrinen K. Topics in social and health informatics research. Finnish Journal of EHealth and EWelfare 2017 May 21;9(2-3):59. [doi: 10.23996/fjhw.63500]
43. Janatzek U. Sozialinformatik – empirisch begründete Zuordnungen und Verständnisweisen. Unter besonderer Berücksichtigung einer wissenschaftstheoretischen Verortung der managerialen Sozialinformatik als Protowissenschaft [Dissertation]. Bielefeld: Universität Bielefeld; 2017.
44. Wulf V, Weibert A, Aal K, Rüller S, Rohde M. The praxeological research programme of socio-informatics - the Siegen School. Acta Inform Pragensia 2021 Dec 31;10(3):333-348. [doi: 10.18267/j.aip.171]
45. Cech F. Exploring emerging topics in social informatics: an online real-time tool for keyword co-occurrence analysis. In: Social Informatics. SocInfo 2017. 2017 Sep 02 Presented at: SocInfo 2017: International Conference on Social Informatics; September 13-15, 2017; Oxford, United Kingdom p. 527-536. [doi: 10.1007/978-3-319-67256-4_42]
46. Hoeffner L. The intellectualsocial structure of social informatics: is there evidence of a domain? [Dissertation]. Brookville, NY: Long Island University; 2012.
47. Meyer E, Shankar K, Willis M, Sharma S, Sawyer S. The social informatics of knowledge. J Assoc Inf Sci Technol 2019 Apr;70(4):307-312 [FREE Full text] [doi: 10.1002/asi.24205] [Medline: 31763357]
48. Friede A, Blum HL, McDonald M. Public health informatics: how information-age technology can strengthen public health. Annu Rev Public Health 1995 May;16:239-252. [doi: 10.1146/annurev.pu.16.050195.001323] [Medline: 7639873]
49. Kharrazi H, Lasser EC, Yasnoff WA, Loonsk J, Advani A, Lehmann HP, et al. A proposed national research and development agenda for population health informatics: summary recommendations from a national expert workshop. J Am Med Inform Assoc 2017 Jan;24(1):2-12 [FREE Full text] [doi: 10.1093/jamia/ocv210] [Medline: 27018264]

## Abbreviations

**HI:** health informatics
**ICT:** information and communication technology
**SI:** social informatics
**WoS:** Web of Science

*Edited by T Leung; submitted 26.06.22; peer-reviewed by A Boyd, Z He; comments to author 02.08.22; revised version received 13.09.22; accepted 21.09.22; published 28.09.22.*

*Please cite as:*
*Smutny Z, Vehovar V*
*On the Current Connection and Relation Between Health Informatics and Social Informatics*
*J Med Internet Res 2022;24(9):e40547*
*URL: https://www.jmir.org/2022/9/e40547*
*doi:10.2196/40547*
*PMID:36169995*

Original Paper

# Psychosocial Mediators of Web-Based Interventions for Promoting a Healthy Lifestyle Among Chinese College Students: Secondary Analysis of a Randomized Controlled Trial

Wei Liang[1,2], PhD; Yanping Duan[1,2], PhD; Yanping Wang[2], MEd; Sonia Lippke[3], PhD; Borui Shang[4], PhD; Zhihua Lin[5], PhD; Hagen Wulff[6], PhD; Julien Steven Baker[1,2], DSc, PhD

[1]Center for Health and Exercise Science Research, Department of Sport, Physical Education and Health, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[2]Department of Sport, Physical Education and Health, Faculty of Social Sciences, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[3]Jacobs University Bremen, Bremen, Germany

[4]Department of Social Sciences, Hebei Sport University, Shijiazhuang, China

[5]Sport Section, Wuhan University, Wuhan, China

[6]Institute of Exercise and Public Health, University of Leipzig, Leipzig, Germany

**Corresponding Author:**
Yanping Duan, PhD
Department of Sport, Physical Education and Health
Faculty of Social Sciences
Hong Kong Baptist University
12/F, Shek Mun Campus
8 On Muk Street, Shek Mun, Shatin
Hong Kong
China (Hong Kong)
Phone: 852 34113038
Fax: 852 34113357
Email: duanyp@hkbu.edu.hk

## Abstract

**Background:** Web-based multiple health behavior change (MHBC) interventions have demonstrated effectiveness in promoting physical activity (PA) and fruit and vegetable consumption (FVC) among Chinese college students. However, there is limited research examining their effects on promoting a healthy lifestyle (ie, adhering to both PA and FVC behavioral recommendations) among Chinese college students. In addition, the salient psychosocial mediators of successful MHBC interventions need to be researched.

**Objective:** This study aims to examine the effectiveness of a previous 8-week web-based MHBC program for promoting a healthy lifestyle and enhancing the psychosocial determinants (intention, self-efficacy, planning, and social support) of behavior change among Chinese college students. Furthermore, the study aims to identify whether changes in these psychosocial determinants mediate intervention effectiveness on the immediate and sustained lifestyle changes.

**Methods:** This was a secondary analysis for a 3-arm randomized controlled trial. Chinese college students (N=552) were randomly assigned to 1 of 3 groups: a PA-first group (4-week PA intervention followed by 4-week FVC intervention), an FVC-first group (4-week FVC intervention followed by 4-week PA intervention), and a placebo control group. The intervention content was designed based on the health action process approach model. Data for analyses were collected at baseline (T0), postintervention assessment (T1), and 12-week follow-up assessment (T2).

**Results:** At baseline, 13.9% (77/552) of the participants maintained a healthy lifestyle. After 8 weeks, more (200/552, 36.2%) participants achieved a healthy lifestyle. PA-first and FVC-first groups were, respectively, 3.24 times and 5 times more likely to adopt a healthy lifestyle than the control group at T1. After 12 weeks, 35.5% (196/552) of the participants adopted a healthy lifestyle. Intervention groups were approximately 2.99 times (PA first) and 4.07 times (FVC first) more likely to adopt a healthy lifestyle than the control group at T2. Intervention effects favored both intervention groups in self-efficacy and planning for PA and in intention and planning for FVC compared with the control condition. In addition, changes in PA self-efficacy and FVC intention mediated intervention effectiveness on the immediate lifestyle change after 8 weeks. Changes in FVC intention were identified as a salient mediator for facilitating sustained lifestyle change after 12 weeks.

**Conclusions:** This study provides empirical evidence for the effectiveness of an 8-week theory- and web-based MHBC intervention program on promoting a healthy lifestyle, self-efficacy and planning for PA, and intention and planning for FVC among Chinese college students. These research findings add new knowledge to the underlying psychosocial mechanisms of successful MHBC interventions. Overall, this study has considerable implications for future web-based MHBC research and practice in terms of addressing PA self-efficacy and FVC intention and helping students to adopt and maintain a healthy lifestyle independently of whether PA or FVC is addressed first.

**Trial Registration:** ClinicalTrials.gov NCT03627949; https://clinicaltrials.gov/ct2/show/NCT03627949

***(J Med Internet Res 2022;24(9):e37563)*** doi:10.2196/37563

**KEYWORDS**

web-based intervention; physical activity; fruit and vegetable consumption; college students; psychosocial mediators; lifestyle; randomized controlled trial; RCT; mobile phone

## *Introduction*

### Background

As common health-protective behaviors, regular physical activity (PA) and adequate fruit and vegetable consumption (FVC) have been proposed as the cornerstones of healthy living [1-3]. These 2 health-protective behaviors constitute a typical pattern of healthy lifestyles, which play a dominant role in improving individuals' overall health [3,4]. An overwhelming body of evidence has demonstrated the considerable impact that increasing PA and FVC could have not only on reducing the morbidity of noncommunicable diseases (eg, cardiovascular diseases, certain types of cancer, gastrointestinal diseases, and obesity) but also on improving mental well-being and quality of life [2,5].

To promote PA and FVC, widely acknowledged behavioral recommendations have been formulated, suggesting that adults aged 18 to 64 years should perform at least 150 minutes of moderate-intensity PA (or at least 75 minutes of vigorous-intensity PA or an equivalent combination of moderate-to-vigorous PA) per week and consume at least five servings (400 g) of fruit and vegetables each day [6,7]. Notwithstanding the recommended behavioral guidelines for PA and FVC, the prevalence of physical inactivity and insufficient intake of fruit and vegetables in the adult population is exceedingly high, especially among college students [8]. In western countries, 23% to 60% of college students do not meet the PA recommendations, whereas <30% comply with the FVC recommendations [8-11]. A similar situation exists in China, where >40% of Chinese college students do not perform the recommended weekly amount of PA, and more than half do not adhere to the recommendation of a minimum 5 servings daily of FVC [12,13]. Therefore, promoting health among college students has become a common challenge in many countries and has stimulated research interest among health psychologists and behavioral scientists [8].

Over the past 4 decades, an increasing number of psychosocial theories have been developed to describe, explain, and predict changes in health behaviors, such as the social cognitive theory [14], health belief model [15], protection motivation theory [16], theory of planned behavior [17], transtheoretical model [18], and the health action process approach (HAPA) [19]. These have been followed by a series of theory-based interventions that seem promising for changing specific health behavior. In particular, interventions based on the HAPA, which integrates the merits of the stage and continuum characteristics of contemporary psychosocial models, have demonstrated remarkable effects on promoting PA, healthy diet, and related health behaviors [20-22].

Although such interventions have achieved singular success, most of them focused only on a specific health behavior and addressed different behaviors as categorically separate entities [23]. As such, the interrelationships among different health behaviors have been artificially disconnected. However, in real life, individuals usually tend to have multidimensional patterns of health behaviors rather than 1 specific behavior in isolation [24]. A key property of these behaviors is that they typically coexist as behavioral clusters or bundles [23-25]. For instance, 1 risk behavior (eg, sedentary behavior) often occurs with other risk behaviors (eg, excessive intake of fat and sugar, smoking, excessive sedentary screen time, or alcohol addiction), or 1 health-protective behavior coexists with other health-protective behaviors (eg, PA and FVC). This high co-occurrence of different health behaviors can generate synergistic or additive effects so that when promoting multiple health-protective behaviors simultaneously, the overall health benefits can be greatly increased [23-27]. As a result, interventions targeting multiple health behavior change (MHBC) have grown in popularity over the past decade as a potential and pragmatic way to maximize overall health outcomes. With the burgeoning use of internet technology, web-based MHBC interventions have been increasingly applied to a wide range of populations [8,28]. Compared with traditional face-to-face hand-delivered interventions this new paradigm has been praised for its numerous advantages, such as accessibility, scalability, cost-effectiveness, flexibility, and convenience [29].

One critical issue that remains understudied concerns the psychosocial mechanisms behind MHBC (ie, salient mediators of successful MHBC interventions) [24,27]. Psychosocial theory-based MHBC interventions have been advocated by many researchers as scientific theories that can provide a useful framework for addressing the key modifiable determinants (eg, motivation and volition) of health behavior that may consequently increase the effectiveness of MHBC interventions [24,27,30,31]. However, many so-called theory-based health interventions are probably better categorized as *theory inspired* rather than *theory based* because they may not apply the theory properly and extensively (eg, the intervention did not effectively link behavior change techniques to the theoretical components) [27,32]. As a result, considerable heterogeneity in the effect sizes is found in theory-based health interventions, and some reviews have even concluded that the use of theories has no bearing on the effectiveness of interventions [27,30-33]. Therefore, it is important to identify active mediators of intervention effectiveness and increase our understanding of theoretical constructs in terms of the magnitude of their impacts in different behavior interventions targeting different populations.

### Our Previous Web-Based MHBC Intervention Program for Chinese College Students

Given the urgent need for, and limited practice of, MHBC interventions for promoting both PA and FVC among Chinese college students, we developed an 8-week web-based MHBC intervention program. To address the debatable question in MHBC research (ie, how to deliver MHBC interventions to achieve more robust treatment effects or whether the order of the sequential intervention contents makes a difference) [34], we designed two sequentially delivered modules (ie, PA first and FVC first) in our previous program and compared the differences in their intervention effects on promoting PA and FVC among Chinese college students (details have been presented elsewhere [34]). Our previous program used the HAPA model as the theoretical backdrop [19]. The HAPA model postulates two distinctive phases of the behavior change process (ie, motivational and volitional phases), underlining the vital role of diverse psychosocial determinants in the behavior change process [35]. In the motivational phase, the primary task is to form a behavioral intention by reinforcing several crucial antecedents (eg, risk perception, outcome expectancies, and action self-efficacy). Once the behavioral intention has been formed, individuals need to enhance maintenance and recovery self-efficacies, apply a series of self-regulatory strategies (eg, action planning and coping planning), and use external resources (eg, social support) for facilitating the behavior initiation and maintenance in the volitional phase [36]. In our previous study, we evaluated the effectiveness of a web-based MHBC intervention program on promoting PA, FVC, and health-related outcomes (BMI, depression, and quality of life) among Chinese college students. The results supported favorable effects on both behaviors and BMI in the intervention groups compared with a control condition, with small-to-medium effect sizes (Cohen $d$=0.22-0.59), and indicated a superior effect on FVC maintenance in the FVC-first group compared with the PA-first group [34]. However, our previous analyses focused only on the change in each specific behavior, whereas the comparative intervention effects on the combination of multiple health behaviors (ie, lifestyle indicator) and psychosocial determinants of behavior change have not been examined. In addition, the underlying psychosocial mechanisms of successful MHBC (salient mediators) have not been identified.

### Objectives and Hypotheses

Given the aforementioned particulars, the first aim of this study was to examine the immediate and sustained effectiveness of our previous 8-week web-based intervention program for promoting a healthy lifestyle (ie, adhering to both PA and FVC behavioral recommendations) and enhancing the psychosocial determinants of PA and FVC (ie, intention, self-efficacy, planning, and social support) among Chinese college students. Furthermore, this study aimed to identify whether changes in psychosocial determinants of PA and FVC could account for the immediate and sustained lifestyle changes (ie, mediation analyses).

Correspondingly, the main intervention effects were hypothesized in terms of greater adoption of a healthy lifestyle (hypothesis 1) and more improvements in the psychosocial determinants of PA (hypothesis 2a) as well as FVC change (hypothesis 2b). The mediation effects were hypothesized in terms of the following assumption: participants in the intervention groups who had increased psychosocial determinants of PA and FVC would be more likely to have positive lifestyle changes after 8 (hypothesis 3a) and 12 weeks (hypothesis 3b) than those in the control group.

## Methods

### Ethics Approval

The study was approved by the research ethics committee of Hong Kong Baptist University (FRG2/15-16/032).

### Design, Participants, and Procedure

The study outlined herein has been described in greater detail elsewhere [34]. Data for the secondary analyses were collected in a 3-arm, double-blinded (ie, intervention facilitator and outcome evaluator) randomized controlled trial (RCT) evaluating sequentially delivered web-based interventions for PA and FVC among Chinese college students (ClinicalTrials.gov NCT03627949) [34].

In our previous RCT, the participants (N=634) were recruited from 28 different departments (the total number of departments is 34) of 1 university in the central region of China. The eligibility criteria were as follows: participants (1) were aged ≥18 years, (2) were not collegiate athletes or had not majored in any sport-related subjects, (3) had no contraindications to physical mobility (eg, cardiovascular diseases and disabilities) or FVC (fruit allergies or diabetes), and (4) had access to the internet and digital devices (eg, desktop computer, laptop computer, and smartphone). Details of the sampling approach, sample size estimate, recruitment procedure, and CONSORT (Consolidated Standards of Reporting Trials) flow diagram are described in detail in the primary paper [34].

After enrollment and eligibility checks, the eligible participants (N=556) were randomly assigned to 1 of 3 groups, which included a PA-first group (4 weeks of PA treatment followed by 4 weeks of FVC treatment), an FVC-first group (4 weeks of FVC treatment followed by 4 weeks of PA treatment), and a control group (8 weeks of placebo treatment irrelevant to either

PA or FVC). The study was implemented from October 2017 to March 2018.

Of the 556 eligible participants, we excluded 4 (0.7%) who did not complete the baseline assessment; thus, the final sample considered for the analysis of this study consisted of 552 (99.3%) participants, specifically 187 (33.9%) in the PA-first group, 195 (35.3%) in the FVC-first group, and 170 (30.8%) in the control group. For the study analyses, measurements were recorded at baseline (T0), after the intervention (T1; 8 weeks after T0), and at follow-up (T2; 12 weeks after T0).

## Intervention

The intervention has been described in greater detail in our previously published papers [34-37]. The intervention content was designed based on the theoretical framework of the HAPA [19], lasting for 8 weeks with 1 session per week (each session lasting for between 20 and 30 minutes). Two sequentially delivered health interventions (ie, PA first and FVC first) were designed to target the HAPA-based psychosocial determinants of PA and FVC change. Considering that >90% of the Chinese college students were intenders and actors for PA and FVC behavior in our previous pilot study [38], this study focused more on the enhancement of intention, self-efficacy, planning, and social support to facilitate the crucial transition from intention to actual behavior initiation and maintenance (ie, intention-behavior gap).

In brief, for the PA-first group, the first 4-week intervention targeted the following psychosocial determinants of PA change:

- Week 1: risk perception, outcome expectancies, and goal settings (these antecedent variables contributed to the formation and enhancement of PA intention)
- Week 2: development of action planning
- Week 3: revision and adjustment of previous action planning and development of coping planning
- Week 4: revision and adjustment of previous coping planning and development of perceived social support

The same intervention materials were subsequently implemented to target the psychosocial determinants of FVC change in the second 4-week intervention period. For the FVC-first group, the sequence of intervention delivery was the reverse of the PA-first module. Self-efficacy was involved as a settled component throughout the entire intervention period. For the control condition, to avoid social desirability and the Hawthorne effect [39], all participants in the control group received active control treatments that seemed in all respects to be identical to the 2 intervention groups (eg, intervention duration, frequency, implementation procedure, and delivery modes) but lacked the critical psychosocial ingredients for changing PA or FVC [34].

The whole intervention, which consisted of three independent modules (ie, 2 MHBC intervention modules and 1 placebo control module), was delivered through a well-established platform. Participants were asked to attend the corresponding intervention session once a week through a laptop computer or desktop computer. WeChat (a popular social media platform in China) groups were established for participants who were included in the same intervention condition. Each participant received a WeChat group reminder that was distributed by the research team 1 day before the new intervention session [34].

## Measures

### Lifestyle Indicator

The lifestyle indicator reflected the combination of multiple health behaviors (ie, whether the participant had complied with behavioral recommendations for both PA and FVC). We used the World Health Organization–recommended thresholds of at least 150 minutes of moderate-intensity PA per week (or at least 75 minutes of vigorous-intensity PA or an equivalent combination of moderate-to-vigorous PA) and 5 daily servings (400 g) of fruit and vegetables [6,7]. In our previous study, the weekly amount of PA was assessed using the Chinese brief version of the International Physical Activity Questionnaire [40], and the daily portion of FVC was evaluated using a Chinese version of the 4-item FVC scale [41]. Participants were categorized into 1 of 2 groups depending on whether they adhered to both PA and FVC recommendations (0=unhealthy lifestyle that met neither of the behavioral recommendations or only 1 behavioral recommendation and 1=healthy lifestyle that met both behavioral recommendations) [42,43].

### Psychosocial Determinants of Behavior Change

#### Intention

Intention for PA was measured with the question stem “I intend to perform at least 30 minutes a day on minimum 5 days a week for at least 150 minutes per week with...” followed by 3 items: “...vigorous PA,” “...moderate PA,” and “...mild PA” (Cronbach $\alpha$=.64). Intention for FVC was assessed by the question stem “I seriously intend to...” followed by 3 items: “...eat at least five servings of fruit and vegetables every day,” “...eat more fruit and vegetables each meal,” and “...drink at least one glass of fruit or vegetable juice every day” (Cronbach $\alpha$=.63). The answers were indicated on a visual analog scale (VAS) ranging from *1=not true* to *4=exactly true* [34,38,43,44].

#### Self-efficacy

Self-efficacy was measured with the question stem “I am certain that...” followed by 5 items for PA such as “...I can be physically active on a permanent and regular basis (eg, at least 30 minutes a day on minimum 5 days a week), even if I have to overcome some barriers” or followed by 5 items for FVC such as “...I can eat 5 portions of fruit and vegetables a day even if it is sometimes difficult” (Cronbach $\alpha$ for PA=.88 and Cronbach $\alpha$ for FVC=.92). The answers were indicated on a VAS ranging from *1=don’t agree at all* to *5=agree completely* [34,38,43-45].

#### Planning

Planning includes two components: action planning and coping planning. Action planning was measured by the question stem “For the next month I already planned in detail...” followed by 3 items for PA such as “...which concrete PA I will pursue” or followed by 3 items for FVC such as “...how I will prepare the food” (Cronbach $\alpha$ for PA=.86 and Cronbach $\alpha$ for FVC=.91). Coping planning was measured by the question stem “For the next month I already planned in detail...” followed by 3 items for PA such as “...how I can stay active, even if something happened” or followed by 3 items for FVC such as “...what I can do in difficult situations, in order to remain true to my own resolutions" (Cronbach α for PA=.87 and Cronbach α for FVC=.93). Answers were given on a VAS ranging from *1=totally disagree* to *5=totally agree* [34,38,43-46].

### Social Support

Perceived social support was measured by the question stem "How do you perceive your environment?" followed by 3 items for PA such as "People like my classmates and friends help me to stay physically active" or followed by 3 items for FVC such as "People like my classmates and friends help me to eat healthily" (Cronbach α for PA=.72 and Cronbach α for FVC=.69). Answers were given on a VAS ranging from *1=disagree* to *4=agree* [34,38,43-46].

### Covariates

The covariates included age, sex, college grade (freshman, sophomore, junior, or senior), marital status (single or in a relationship), perceived health status (poor, satisfactory, or excellent), and BMI (kg/m$^2$) [47].

All the questionnaires were written in simple Chinese and had been validated in previous studies using Chinese adult populations [34,38,43]. Sociodemographic information was collected only at registration, whereas all other indicators were assessed at baseline (T0), postintervention assessment (T1), and 12-week follow-up assessment (T2).

## *Statistical Analyses*

Data analyses were performed using SPSS software (version 27.0; IBM Corp; eg, descriptive tests and intervention effect evaluation) and PROCESS macro (version 4.0; Andrew F Hayes; mediation analyses). Baseline characteristics and randomization were checked using independent 2-tailed *t* tests, ANOVA, and chi-square tests. Missing values were imputed using the multiple imputation approach with chained equations, except for dropouts, which were addressed using the baseline-observation-carried-forward approach [48]. The 5% level (2-tailed) was used as the statistical significance cutoff point.

With an intention-to-treat principle, intervention effects on the lifestyle indicator were examined using logistic regression analyses (determining odds ratios; hypothesis 1). For intervention effects on psychosocial determinants of behavior change (hypotheses 2a and 2b), generalized linear mixed models were used using a restricted maximum likelihood approach with time, group, and their interaction as fixed effects adjusted for the random effects of baseline behaviors. Unstructured covariance matrix was selected based on the minimal values of −2 log likelihood and Akaike and Bayesian information criteria. The least significant difference method was used for the post hoc comparison [49].

For hypotheses 3a and 3b, to control for the effects of baseline values, residualized change scores were used for the multiple mediation analyses [50]. The standardized coefficients and 95% CIs for direct, indirect, and total effects were estimated using the bias-corrected bootstrap approach (5000 resamples). The multicollinearity of psychosocial mediators was checked before the mediation analyses using the following criteria for an ignorable multicollinearity problem: low correlation (≤0.70), high tolerance (>0.01), low variance inflation factor (≤10), high eigenvalue (not approaching 0), and small condition index (≤30) [51]. For effect size $R^2$, the proposed small, medium, and large values were 0.02, 0.13, and 0.26, respectively [52].

# Results

### Sample Characteristics and Randomization Check

A total of 552 participants (n=322, 58.3%, women) were included in the data analysis, with their ages ranging from 18 to 24 (mean 19.99, SD 1.04) years. Table 1 presents the descriptive information of the study sample in terms of their sociodemographic data, baseline values of psychosocial determinants for PA and FVC, and behavioral indicators at baseline.

Randomization checks indicated that there were no significant differences in baseline characteristics across the 3 groups in relation to age, sex, college grade, marital status, perceived health status, and BMI (*P*=.37-.83). In addition, the 3 groups did not vary significantly in all psychosocial mediators and behavioral indicators (*P*=.10-.93). Therefore, the randomization was successful.

At baseline (T0), 27.9% (154/552) of the participants did not meet the weekly PA recommendation, whereas 80.4% (444/552) did not consume at least five portions of fruit and vegetables per day. When both behaviors were combined, 86.1% (475/552) of the participants met only 1 or none of these 2 behavioral recommendations and were categorized as adopting unhealthy lifestyles at baseline. Overall, 13.9% (77/552) of the participants achieved both behavioral recommendations and were categorized as adopting healthy lifestyles.

At T1, 23.2% (128/552) of the participants did not meet the PA recommendation, whereas 55.6% (207/552) did not achieve the recommended daily servings of fruit and vegetables. When both behaviors were combined, 36.2% (200/552) of the participants met both behavioral recommendations and were categorized as having healthy lifestyles.

At T2, the percentage of participants adhering to the PA recommendation was 21.2% (117/552), whereas 43.8% (242/552) met the FVC recommendation. Taking both behaviors together, 35.5% (196/552) of the participants complied with both PA and FVC recommendations and were categorized as adopting healthy lifestyles.

**Table 1.** Sociodemographic information, psychosocial mediators, and behavioral indicators of the study sample at baseline.

| Variable | Total (N=552) | PA[a]-first group (n=187) | FVC[b]-first group (n=195) | Control group (n=170) |
|---|---|---|---|---|
| **Sociodemographic information** | | | | |
| Age (range 18-24 years), mean (SD) | 19.99 (1.04) | 20.07 (1.07) | 19.96 (0.99) | 19.93 (1.06) |
| **Sex, n (%)** | | | | |
| Male | 230 (41.7) | 79 (42.2) | 78 (40) | 73 (42.9) |
| Female | 322 (58.3) | 108 (57.8) | 117 (60) | 97 (57.1) |
| **College grade, n (%)** | | | | |
| Freshman | 264 (47.8) | 86 (46) | 90 (46.2) | 88 (51.8) |
| Sophomore | 229 (41.5) | 77 (41.2) | 84 (43.1) | 68 (40) |
| Junior | 46 (8.3) | 18 (9.6) | 16 (8.2) | 12 (7.1) |
| Senior | 13 (2.4) | 6 (3.2) | 5 (2.6) | 2 (1.2) |
| **Marital status, n (%)** | | | | |
| Single | 506 (91.7) | 170 (90.9) | 183 (93.8) | 153 (90) |
| In a relationship | 46 (8.3) | 17 (9.1) | 12 (6.2) | 17 (10) |
| **Perceived health status, n (%)** | | | | |
| Poor | 17 (3) | 5 (2.7) | 9 (4.6) | 3 (1.8) |
| Satisfactory | 358 (64.9) | 122 (65.2) | 125 (64.1) | 111 (65.3) |
| Excellent | 177 (32.1) | 60 (32.1) | 61 (31.3) | 56 (32.9) |
| BMI (range 15.62-32.88 kg/m$^2$), mean (SD) | 20.41 (2.45) | 20.32 (2.34) | 20.52 (2.62) | 20.40 (2.39) |
| **Psychosocial determinants, mean (SD)** | | | | |
| PA intention | 2.22 (0.71) | 2.26 (0.74) | 2.22 (0.72) | 2.17 (0.69) |
| PA self-efficacy | 2.96 (1.19) | 3.08 (1.22) | 2.87 (1.16) | 2.92 (1.2) |
| PA planning | 3.03 (1.05) | 3.10 (0.98) | 2.97 (1.04) | 3.04 (1.12) |
| PA social support | 2.23 (0.91) | 2.25 (0.90) | 2.24 (0.94) | 2.19 (0.9) |
| FVC intention | 1.96 (0.79) | 1.93 (0.76) | 1.97 (0.81) | 1.99 (0.81) |
| FVC self-efficacy | 3.08 (1.37) | 3.07 (1.34) | 3.06 (1.36) | 3.11 (1.42) |
| FVC planning | 2.86 (1.16) | 2.83 (1.16) | 2.83 (1.14) | 2.94 (1.18) |
| FVC social support | 2.37 (0.86) | 2.40 (0.83) | 2.38 (0.89) | 2.32 (0.86) |
| **Behavioral indicators, mean (SD)** | | | | |
| PA (minutes per week) | 465.85 (257.29) | 482.63 (269.42) | 452.80 (248.94) | 462.38 (253.51) |
| FVC (portions per day) | 3.81 (1.75) | 3.84 (1.70) | 3.82 (1.87) | 3.76 (1.68) |
| **Lifestyle indicator, n (%)** | | | | |
| Unhealthy[c] | 475 (86.1) | 161 (86.1) | 166 (85.1) | 148 (87.1) |
| Healthy[d] | 77 (13.9) | 26 (13.9) | 29 (14.9) | 22 (12.9) |

[a] PA: physical activity.

[b] FVC: fruit and vegetable consumption.

[c] Participants adopted unhealthy lifestyles that met neither of the behavioral recommendations or only 1 behavioral recommendation.

[d] Participants adopted healthy lifestyles that met both behavioral recommendations.

## Intervention Effects on Lifestyle Indicator

After the 8-week intervention (T1), both intervention groups, particularly the PA-first group, outperformed the control group in adhering to both PA and FVC behavioral recommendations (39% vs 18.2%). At the 1-month follow-up test (T2), the favorable effects were sustained for both intervention groups: 39% (73/187) of the participants in the PA-first group and 46.2% (90/195) of those in the FVC-first group adopted a healthy lifestyle, whereas only 19.4% (33/170) of the participants in the control condition did so (Figure 1).

To further explore the extent to which the intervention predicted the adoption of a healthy lifestyle at T1 and T2, binary logistic regression analyses were used (Table 2). First, all sociodemographic variables and intervention groups were used as predictors for the adoption of a healthy lifestyle at baseline. Neither of these variables showed a significant correlation to the lifestyle indicator (all *P*=.23-.49). When controlling for all sociodemographic variables and baseline lifestyle, the treatment was found to be a significant predictor for adopting a healthy lifestyle at both T1 and T2 (all *P*<.001). Specifically, after 8 weeks, participants in the PA-first and FVC-first groups were approximately 3.2 times and 5 times more likely, respectively, to practice or maintain a healthy lifestyle than those in the control group. After 12 weeks, participants receiving the interventions were approximately 3 times (PA first) and 4.1 times (FVC first) more likely to comply with a healthy lifestyle than the control group. The entire model accounted for 22% and 21% of the variance of the lifestyle indicator at T1 and T2, respectively.

**Figure 1.** Descriptive information of lifestyle indicator from baseline assessment (T0) to follow-up assessment 12 weeks after baseline assessment (T2). FVC: fruit and vegetable consumption; PA: physical activity; T1: postintervention assessment 8 weeks after baseline assessment.

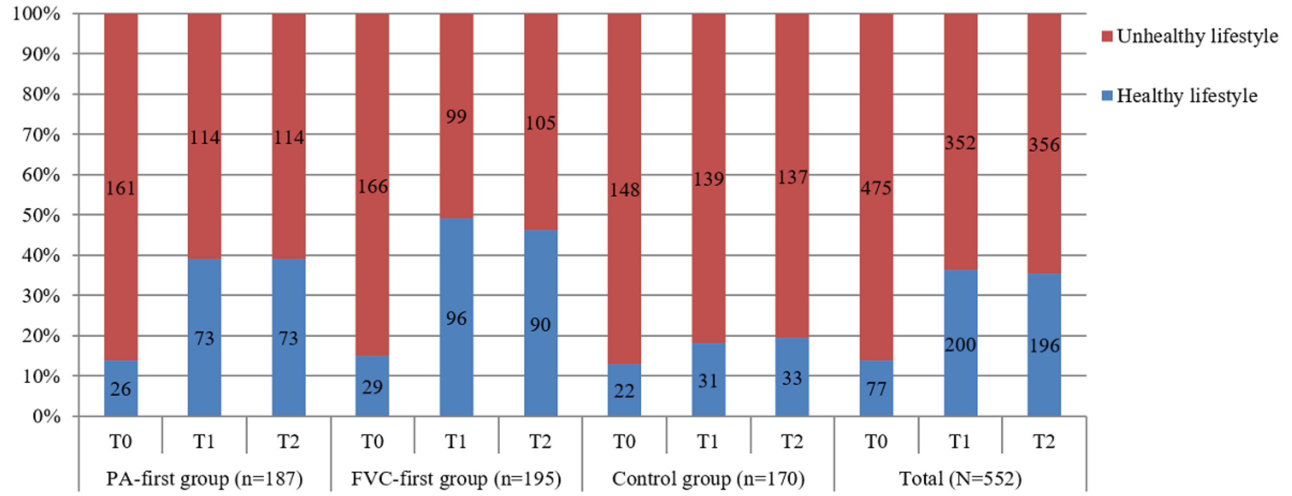


**Table 2.** Intervention effects on the adoption of healthy lifestyles after 8 and 12 weeks (N=552).

| Variable | Lifestyle[a] at baseline (T0) | Lifestyle[a] after 8 weeks (T1) | Lifestyle[a] after 12 weeks (T2) |
|---|---|---|---|
| Constant | <.001 | 0.30 | 0.10 |
| Lifestyle[a] at baseline, OR[b] (95% CI) | N/A[c] | 0.14 (0.08-0.25)[d] | 0.13 (0.07-0.23)[d] |
| PA[e]-first group, OR (95% CI) | 1.10 (0.60-2.03) | 3.24 (1.92-5.47)[d] | 2.99 (1.78-5.03)[d] |
| FVC[f]-first group, OR (95% CI) | 1.19 (0.65-2.17) | 5.00 (2.98-8.40)[d] | 4.07 (2.44-6.80)[d] |
| *R* | 0.02 | 0.22 | 0.21 |

[a] 0=adopted an unhealthy lifestyle (as reference); 1=adopted a healthy lifestyle.

[b] OR: odds ratio, adjusted for all sociodemographic variables.

[c] N/A: not applicable.

[d] *P*<.001.

[e] PA: physical activity (control group was set as reference).

[f] FVC: fruit and vegetable consumption (control group was set as reference).

## Intervention Effects on Psychosocial Determinants of Behavior Change

The results of the linear mixed models showed that of the 8 *time* and *group* interactions, 4 (50%) were statistically significant (Tables 3 and 4). The marginal mean values of the psychosocial determinants of PA and FVC at 3 time points are presented in Figures 2 and 3.

Regarding the psychosocial determinants of PA, the interaction effect of time and treatment on self-efficacy (*P*<.001) and planning (*P*=.008) was significant for both intervention groups compared with the control group. After 8 weeks (T1), a significant between-group difference was found on intention (*P*<.001), self-efficacy (*P*=.01), and planning (*P*=.003-.008), with small-to-medium effect sizes (Cohen *d*=0.26-0.39), which was in favor of the 2 intervention groups. After 12 weeks (T2), the 2 intervention groups showed superiority in the improvement in all psychosocial determinants of PA (Cohen *d*=0.23-0.45), except for a nonsignificant difference in the perceived social support between the FVC-first and control groups (*P*=.16). Two intervention groups did not show any significant differences in the post hoc comparison (*P*=.43 at T1 and *P*=.93 at T2).

For psychosocial determinants of FVC, a statistically significant interaction effect was found on intention (*P*<.001) and planning (*P*<.001), whereas the time×group effect was marginally significant for self-efficacy (*P*=.06) and nonsignificant for social support (*P*=.83). After 8 weeks (T1), a significant between-group difference was found only in FVC intention (*P*<.001; Cohen *d*=0.39-0.45), which favored the 2 intervention groups. There were no significant between-group differences in other variables (*P*=.07-.67). After 12 weeks (T2), the 2 intervention groups showed a prominently higher level of intention (*P*<.001) and planning for FVC (*P*=.002-.04) than the control group. In addition, a significant difference in FVC self-efficacy was found between the FVC-first and control groups (*P*=.02), whereas the PA-first group showed a favorable change in FVC social support compared with the control group (*P*=.046).

**Table 3.** Results of the generalized linear mixed models with psychosocial mediators of physical activity (PA) change after 8 and 12 weeks as outcome measures (N=552).

| Time and group | PA intention | | PA self-efficacy | | PA planning | | PA social support | |
|---|---|---|---|---|---|---|---|---|
| | Value | Effect size, Cohen *d* | Value | Effect size, Cohen *d* | Value | Effect size, Cohen *d* | Value | Effect size, Cohen *d* |
| **Type III tests, *F*[a]** | | | | | | | | |
| Time×group | 2.189 | N/A[b] | 5.55[c] | N/A | 3.49[d] | N/A | 0.83 | N/A |
| Time | 5.88[d] | N/A | 1.10 | N/A | 1.96 | N/A | 6.76[d] | N/A |
| Group | 5.49[d] | N/A | 5.34[d] | N/A | 3.73[e] | N/A | 1.93 | N/A |
| **After 8 weeks (T1), difference of marginal means[f]** | | | | | | | | |
| PA-first group versus control | 0.28[c] | 0.39 | 0.31[e] | 0.27 | 0.34[d] | 0.32 | 0.18 | 0.20 |
| FVC[g]-first group versus control | 0.27[c] | 0.38 | 0.30[e] | 0.26 | 0.30[d] | 0.30 | 0.17 | 0.19 |
| PA-first group versus FVC-first group | 0.003 | 0.004 | 0.01 | 0.01 | 0.04 | 0.04 | 0.01 | 0.01 |
| **After 12 weeks (T2), difference of marginal means[f]** | | | | | | | | |
| PA-first group versus control | 0.30[d] | 0.33 | 0.50[c] | 0.43 | 0.37[d] | 0.34 | 0.21[e] | 0.23 |
| FVC-first group versus control | 0.24[e] | 0.26 | 0.52[d] | 0.45 | 0.34[d] | 0.31 | 0.13 | 0.15 |
| PA-first group versus FVC-first group | 0.06 | 0.07 | –0.02 | –0.02 | 0.03 | 0.03 | 0.07 | 0.08 |

[a] Adjusted for baseline physical activity (metabolic equivalent minutes per week).

[b] N/A: not applicable.

[c] *P*<.001.

[d] *P*<.01.

[e] *P*<.05.

[f] Post hoc test: least significant difference.

[g] FVC: fruit and vegetable consumption.

**Table 4.** Results of the generalized linear mixed models with psychosocial mediators of fruit and vegetable consumption (FVC) change after 8 and 12 weeks as outcome measures (N=552).

| Time and group | FVC intention | | FVC self-efficacy | | FVC planning | | FVC social support | |
|---|---|---|---|---|---|---|---|---|
| | Value | Effect size, Cohen *d* | Value | Effect size, Cohen *d* | Value | Effect size, Cohen *d* | Value | Effect size, Cohen *d* |
| **Type III tests, *F*[a]** | | | | | | | | |
| Time×group | 6.66[b] | N/A[c] | 2.28 | N/A | 5.41[b] | N/A | 0.37 | N/A |
| Time | 39.06[b] | N/A | 6.98[d] | N/A | 17.62[b] | N/A | 5.73[d] | N/A |
| Group | 5.90[d] | N/A | 0.94 | N/A | 0.96 | N/A | 2.09 | N/A |
| **After 8 weeks (T1), mean difference[e]** | | | | | | | | |
| PA[f]-first group versus control | 0.33[b] | 0.39 | 0.14 | 0.11 | 0.19 | 0.16 | 0.17 | 0.20 |
| FVC-first group versus control | 0.39[b] | 0.45 | 0.22 | 0.18 | 0.23 | 0.20 | 0.17 | 0.19 |
| PA-first group versus FVC-first group | –0.06 | –0.07 | –0.08 | –0.07 | –0.05 | –0.04 | 0.01 | 0.01 |
| **After 12 weeks (T2), mean difference[e]** | | | | | | | | |
| PA-first group versus control | 0.32[b] | 0.35 | 0.22 | 0.17 | 0.24[g] | 0.22 | 0.18[g] | 0.21 |
| FVC-first group versus control | 0.37[b] | 0.41 | 0.32[g] | 0.25 | 0.35[e] | 0.32 | 0.15 | 0.17 |
| PA-first group versus FVC-first group | –0.05 | –0.06 | –0.10 | –0.08 | –0.11 | –0.10 | 0.03 | 0.04 |

[a] Adjusted for baseline fruit and vegetable consumption (portions per day).

[b] *P*<.001.

[c] N/A: not applicable.

[d] *P*<.01.

[e] Post hoc test: least significant difference.

[f] PA: physical activity.

[g] *P*<.05.

**Figure 2.** Marginal mean values of psychosocial determinants of physical activity (PA) change for 3 groups from baseline assessment (T0) to follow-up assessment 12 weeks after baseline assessment (T2). (A) Intention for PA. (B) Self-efficacy for PA. (C) Planning for PA. (D) Social support for PA. FVC: fruit and vegetable consumption; T1: postintervention assessment 8 weeks after baseline assessment.

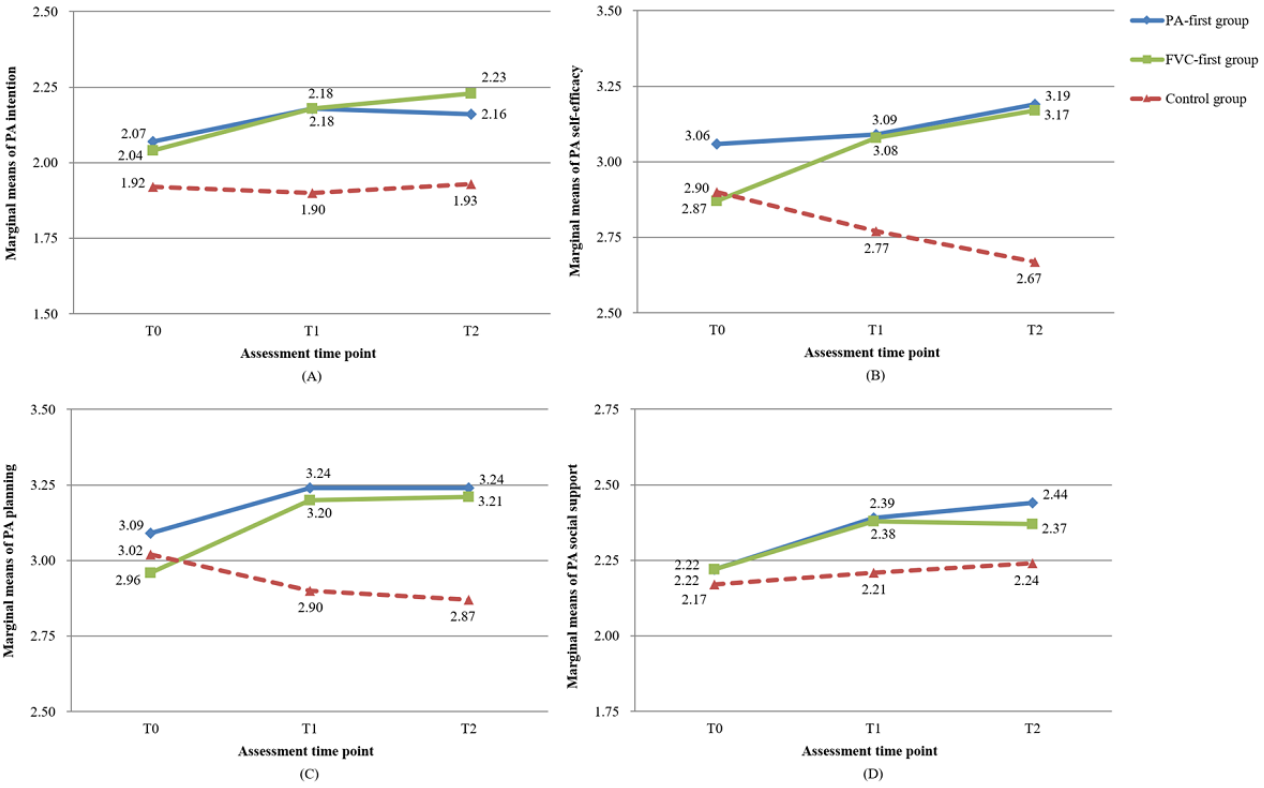
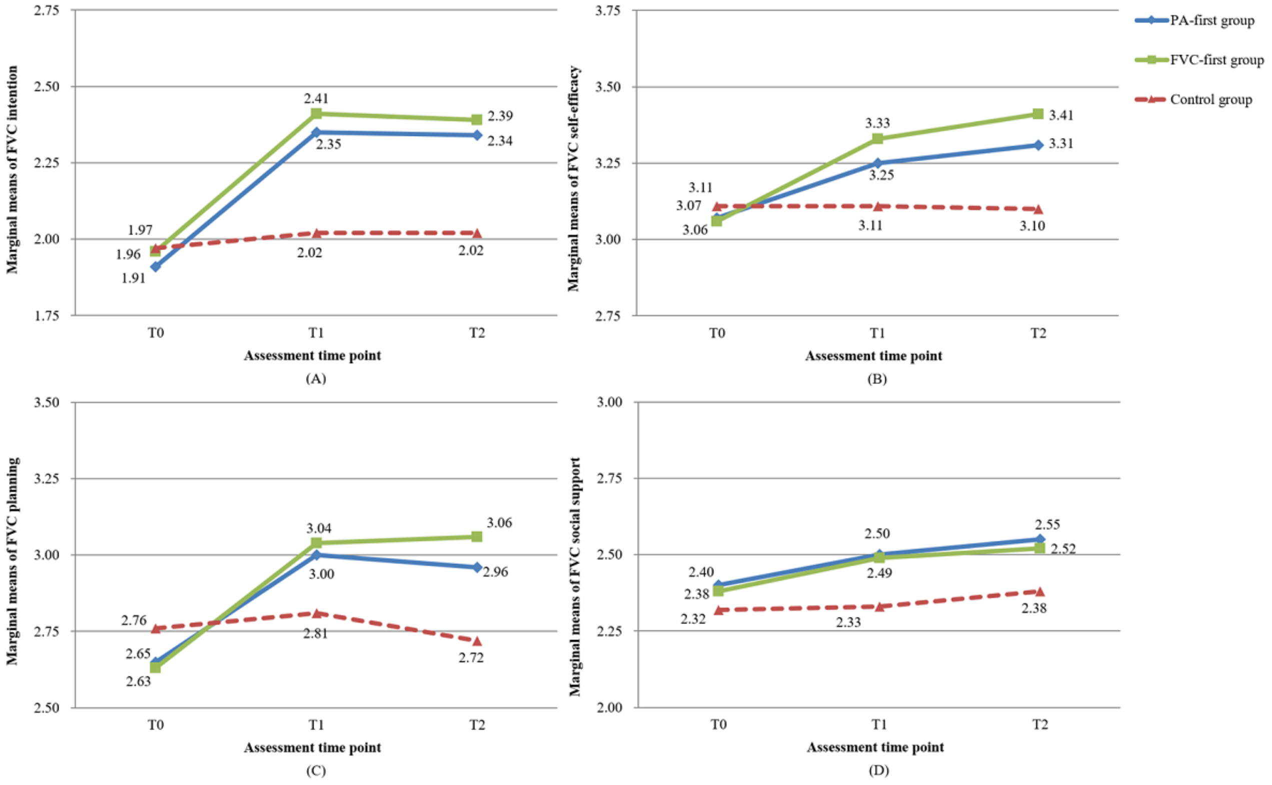


**Figure 3.** Marginal mean values of psychosocial determinants of fruit and vegetable consumption (FVC) change for 3 groups from baseline assessment (T0) to follow-up assessment 12 weeks after baseline assessment (T2). (A) Intention for FVC. (B) Self-efficacy for FVC. (C) Planning for FVC. (D) Social support for FVC. PA: physical activity; T1: postintervention assessment 8 weeks after baseline assessment.



## Mediation Mechanisms of Immediate and Sustained Lifestyle Changes

Multicollinearity diagnostics revealed that there were no severe collinearity problems among the included psychosocial determinants of PA and FVC (correlation $r$=0.33-0.59, tolerance=0.45-0.72, variance inflation factor=1.40-2.25, eigenvalue=0.30-2.67, and condition index=1.00-2.93). Residualized change scores were obtained from the linear or binary regression analyses of T1 scores on T0 scores (ie, immediate change after 8 weeks) and of T2 scores on T0 scores (ie, sustained change after 12 weeks). All sociodemographic variables were included as covariates in the mediation analyses.

After 8 weeks (T1), both intervention assignments significantly predicted the lifestyle changes ($b_{\text{PA first}}$=0.45, 95% CI 0.25-0.65; $P$<.001; $b_{\text{FVC first}}$=0.66, 95% CI 0.21-0.62; $P$<.001) and changes in all psychosocial determinants of behavior change, except social support (Figure 4). After controlling for the changes in psychosocial determinants, the associations between group assignments and lifestyle changes were attenuated but still statistically significant ($b_{\text{PA first}}$=0.31, 95% CI 0.12-0.51; $P$=.002; $b_{\text{FVC first}}$=0.51, 95% CI 0.31-0.70; $P$<.001), indicating that PA self-efficacy and FVC intention were partial mediators of intervention effectiveness. The multiple mediator model accounted for 17.5% of the variance in immediate lifestyle changes ($P$<.001).

After 12 weeks (T2), the intervention assignments continuously showed a significant prediction for lifestyle changes ($b_{\text{PA first}}$=0.42, 95% CI 0.22-0.63; $P$<.001; $b_{\text{FVC first}}$=0.57, 95% CI 0.37-0.78; $P$<.001) and changes in all psychosocial mediators, except social support (Figure 5). Among 8 mediators, only FVC intention was identified as a significant mediator that partially mediated the effects of both intervention groups on lifestyle changes at T2 ($b$=0.27, 95% CI 0.17-0.37; $P$<.001). The overall mediation model accounted for 18.4% of the variance in sustained lifestyle changes ($P$<.001).

**Figure 4.** Mediation model of intervention effects on immediate lifestyle change at T1 (postintervention assessment 8 weeks after baseline assessment). CG: control group; FVC: fruit and vegetable consumption; PA: physical activity. *$P$<.05, **$P$<.01, ***$P$<.001.

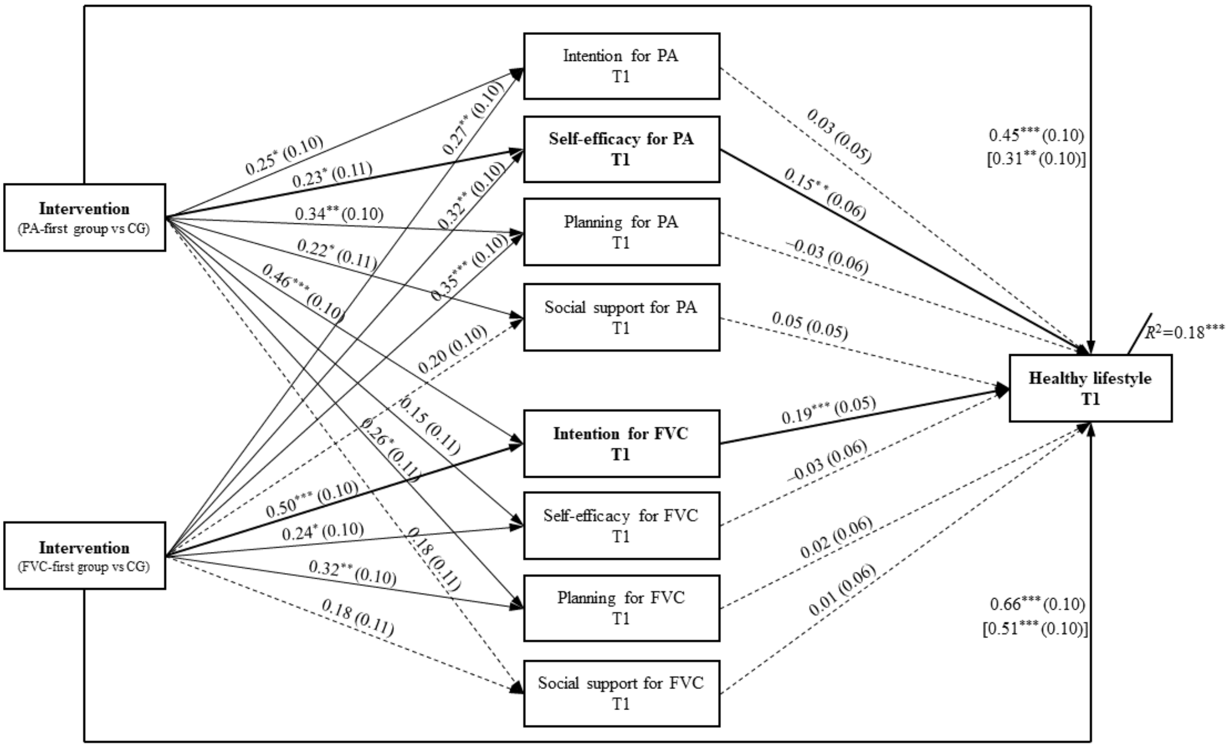
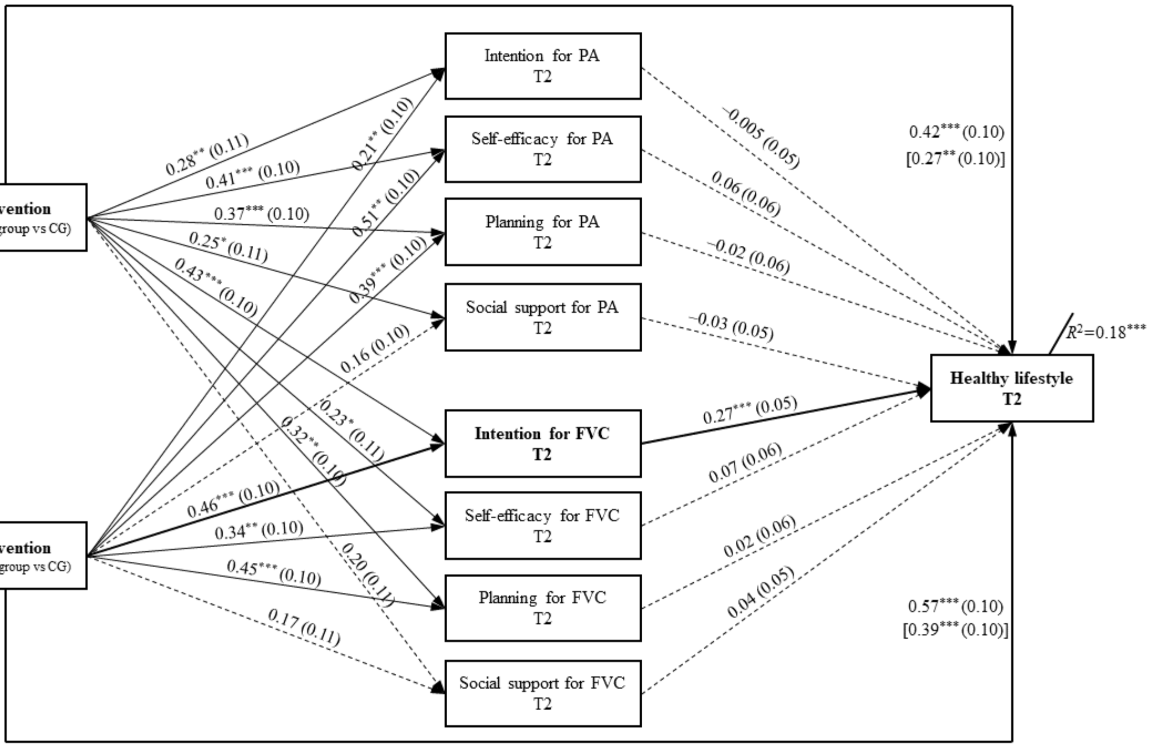


**Figure 5.** Mediation model of intervention effects on sustained lifestyle change at T2 (follow-up assessment 12 weeks after baseline assessment). CG: control group; FVC: fruit and vegetable consumption; PA: physical activity. *$P<.05$, **$P<.01$, ***$P<.001$.



# Discussion

## Principal Findings

This is one of the first studies to gain insights into web-based MHBC interventions for Chinese college students in terms of (1) the intervention effects on promoting a healthy lifestyle and enhancing psychosocial determinants of behavior change and (2) the mediation mechanisms of immediate and sustained lifestyle changes. The majority of the study hypotheses were supported.

The principal expected intervention effects on prompting a healthy lifestyle were identified (hypothesis 1). Compared with the participants in the placebo control condition, those in the intervention groups showed immediate and sustained improvements in adhering to both PA and FVC behavioral recommendations after receiving the 8-week web-based MHBC interventions, with medium effect sizes ($R^2$=0.21-0.22). This finding is consistent with a previous computer-based intervention study with German employees [42,53], Iranian adolescents [20], and the general population in different countries [8,29]. Similar results were also presented in our previous study that aimed to enhance PA and FVC among outpatients with coronary heart disease during their home-based rehabilitation [43]. Taken together, our findings add evidence regarding the potential of web-based MHBC interventions for promoting a healthy lifestyle among young adult populations.

Regarding the intervention effects on the psychosocial determinants of behavior change, 50% (4/8) were found to be statistically significant. Research hypotheses 2a and 2b were partially supported. Although a favorable improvement was detected for the 2 intervention groups descriptively, we could not find a statistically significant time and treatment interaction on intention for PA, self-efficacy for FVC, and social support for both PA and FVC. The findings were inconsistent with those of previous studies of college students and other populations that had indicated a significant intervention effect on these variables [38]. One potential interpretation could be that the ceiling effect came into play here [54]. In particular, the participants in this study had a high level of intention for PA (mean 2.22, SD 0.71; scale scoring range 1-4) and high perceived social support for both health behaviors (mean$_{PA}$ 2.23, SD 0.91; mean$_{FVC}$ 2.37, SD 0.86) at baseline. In addition, our findings might be attributed to the impacts of external social and environmental factors (eg, university policy and environmental barriers). In our previous qualitative interviews, these participants had stated that their health behaviors are considerably affected by the mandatory university policy for PA in terms of the *Ham Run task* (ie, all undergraduates had to complete a 2000-meter run 28 times, accounting for 20% of the PE course credit) and barriers to FVC (eg, financial issues and limited provision of fruit and vegetables at university canteens) [34]. Unsurprisingly, in such a case these external sources might, to some extent, suppress the intervention effects on the internal sources of behavior change (eg, intention for PA and self-efficacy for FVC). As our research focused on individual-level psychosocial determinants of behavior change, the social and environmental factors were not involved. This should be systematically examined in future studies.

For mediation analyses (hypotheses 3a and 3b), only intention and self-efficacy were identified as salient mediators of lifestyle changes. In particular, compared with the control condition participants, those in the intervention groups who gained more self-efficacy for PA and who increased more intention for FVC were more likely to show a successful change in lifestyle after 8 weeks (ie, immediate change). This finding is consistent with that of previous studies of workplace employees and clinical patients [42,43,53], demonstrating the importance of empowering the internal sources (intention and self-efficacy) in facilitating both sufficient PA and healthy diets. For the sustained lifestyle change (after 12 weeks), only intention for FVC was identified as a significant mediator of intervention effectiveness. This finding supplements the evidence for emphasizing the role of intention in maintaining long-term change of lifestyle behaviors [55,56]. The hypothesized role of planning and social support in facilitating a healthy lifestyle has not been found in our study, and further investigation is warranted. In addition, the mediation models only showed medium effect sizes in explaining the variance of lifestyle changes among Chinese college students ($R^2$=0.18), which are comparatively higher than those of German workplace employees ($R^2$=0.10) [42] and lower than those of Chinese outpatients with coronary heart disease ($R^2$=0.33) [43]. Further studies with inclusion of more psychosocial mediators are warranted.

## Limitations

Several limitations should be noted. First, the behavioral indicators were evaluated using self-reported measures, which may lead to recall bias and social desirability effect [57]. The inclusion of objective measures such as accelerometers, pedometers, and digital cameras, which can provide more accurate and reliable assessments of health behaviors, is recommended in future studies. Second, the RCT design was used in consideration of the feasibility and limited resources for study implementation; however, this may lead to spillover effect and contamination [58]. Although we applied several strategies to minimize this problem and did not identify any contamination in our previous study, a stricter design (eg, cluster RCT) should be used, if possible. In addition, the intervention effects may be confounded by external sources (eg, season, university policy, PA facilities, and environmental barriers) [59]. Further investigation considering these factors is warranted. In addition, following a parsimonious principle and considering the characteristics of the study sample, we did not include habit strength and action control as in our previous intervention program. Accordingly, the role of these factors in facilitating a healthy lifestyle has not been examined in this study. However, further identification of the mediating effect of these factors is needed [60]. In addition, because this is a secondary analysis of our previous RCT targeting specific outcomes (healthy lifestyle as well as psychosocial determinants of PA and FVC), the findings generated in this study cannot be regarded as representative of all student samples who receive the web-based MHBC intervention, and caution is needed when generalizing to wider populations. Finally, our study focused on the lifestyle pattern combining only PA and FVC; more lifestyle behaviors (eg, sedentary behaviors, sleep patterns, smoking, and alcohol addiction) are deserving of inclusion in future studies to contribute to a better understanding of comprehensive lifestyle patterns. Despite these limitations, this study may have considerable implications for future MHBC research and practice on promoting a healthy lifestyle among college students in terms of addressing PA self-efficacy and FVC intention. Our findings supplement evidence on the effectiveness of web-based MHBC interventions independently of whether PA or FVC is targeted first. The study adds new knowledge about the underlying mechanisms of successful MHBC interventions in terms of lifestyle approaches that require combined strategies.

## Conclusions

To conclude, this study demonstrated the great potential of 8-week theory-based and web-based MHBC interventions for promoting a healthy lifestyle and several psychosocial determinants of behavior change among Chinese college students. This study also identified a salient mediating effect of intention and self-efficacy in facilitating successful, immediate, or sustained lifestyle changes. The research findings provide empirical evidence for future MHBC research and practice among young adult populations: lifestyle can be improved independently of whether PA or FVC is addressed first by means of web-based interventions. Further investigation on the effects in other populations and countries and with other behaviors, such as healthy internet use and stress management, is needed.

## Acknowledgments

This research was funded by the Faculty Research Grant of Hong Kong Baptist University (FRG2/15-16/032; principal investigator: YD) as well as WL's Doctoral Dissertation Fellowship through Hong Kong Baptist University. The funding organization had no role in the study design, study implementation, manuscript preparation, or publication decision. This work is the responsibility of the authors.

## Authors' Contributions

WL was responsible for conceptualization, investigation, methodology, data collection and analysis, and writing (original draft, revision, and editing). YD was responsible for conceptualization, funding acquisition, supervision, investigation, methodology, and writing (revision and editing). YW was responsible for investigation and methodology. SL was responsible for conceptualization and writing (revision). ZL was responsible for investigation and data collection. HW and JSB were responsible for revision and language polishing. All authors reviewed and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

## References

1. de Ridder D, Kroese F, Evers C, Adriaanse M, Gillebaart M. Healthy diet: health impact, prevalence, correlates, and interventions. Psychol Health 2017 Aug;32(8):907-941. [doi: 10.1080/08870446.2017.1316849] [Medline: 28447854]
2. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
3. Lachat C, Otchere S, Roberfroid D, Abdulai A, Seret FM, Milesevic J, et al. Diet and physical activity for the prevention of noncommunicable diseases in low- and middle-income countries: a systematic policy review. PLoS Med 2013;10(6):e1001465 [FREE Full text] [doi: 10.1371/journal.pmed.1001465] [Medline: 23776415]
4. Prochaska JJ, Spring B, Nigg CR. Multiple health behavior change research: an introduction and overview. Prev Med 2008 Mar;46(3):181-188 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.001] [Medline: 18319098]
5. Wang X, Ouyang Y, Liu J, Zhu M, Zhao G, Bao W, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 2014 Jul 29;349:g4490 [FREE Full text] [doi: 10.1136/bmj.g4490] [Medline: 25073782]
6. WHO guidelines for physical activity and sedentary behavior. World Health Organization. Geneva, Switzerland: World Health Organization; 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/336657/9789240015111-eng.pdf [accessed 2022-08-11]
7. Milton K, Cavill N, Chalkley A, Foster C, Gomersall S, Hagstromer M, et al. Eight investments that work for physical activity. J Phys Act Health 2021 Jun 01;18(6):625-630. [doi: 10.1123/jpah.2021-0112] [Medline: 33984836]
8. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. Prev Med 2017 Jun;99:197-206. [doi: 10.1016/j.ypmed.2017.01.009] [Medline: 28130046]
9. Aceijas C, Waldhäusl S, Lambert N, Cassar S, Bello-Corassa R. Determinants of health-related lifestyles among university students. Perspect Public Health 2017 Jul;137(4):227-236. [doi: 10.1177/1757913916666875] [Medline: 27597797]
10. Kyrkou C, Tsakoumaki F, Fotiou M, Dimitropoulou A, Symeonidou M, Menexes G, et al. Changing trends in nutritional behavior among university students in Greece, between 2006 and 2016. Nutrients 2018 Jan 10;10(1):64 [FREE Full text] [doi: 10.3390/nu10010064] [Medline: 29320449]
11. Pengpid S, Peltzer K, Kassean HK, Tsala Tsala JP, Sychareun V, Müller-Riemenschneider F. Physical inactivity and associated factors among university students in 23 low-, middle- and high-income countries. Int J Public Health 2015 Jul;60(5):539-549. [doi: 10.1007/s00038-015-0680-0] [Medline: 25926342]
12. Dong H, Wang Y. Investigation of dietary behavior and physical activity in university students. China Health Care Nutr 2018;28:242-243 [FREE Full text] [doi: 10.3969/j.issn.1004-7484.2018.34.329]
13. Li X, Yang H, Yang F. Influences of campus environment on physical activity participation of college students. J Wuhan Inst Phys Educ 2018;52:74-81. [doi: 10.15930/j.cnki.wtxb.2018.01.012]
14. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001 Feb;52(1):1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
15. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
16. Rogers RW. Cognitive and psychological processes in fear appeals and attitude change: a revised theory of protection motivation. In: Cacioppo JT, Petty RE, editors. Social Psychophysiology: A Sourcebook. New York, NY, USA: The Guilford Press; 1983:153-176.
17. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control: From Cognition to Behavior. Heidelberg, Germany: Springer; 1985:11-39.
18. Prochaska JO, DiClemente CC. Transtheoretical therapy: toward a more integrative model of change. Psychother 1982;19(3):276-288. [doi: 10.1037/h0088437]
19. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
20. Lin CY, Scheerman JF, Yaseri M, Pakpour AH, Webb TL. A cluster randomised controlled trial of an intervention based on the Health Action Process Approach for increasing fruit and vegetable consumption in Iranian adolescents. Psychol Health 2017 Dec;32(12):1449-1468. [doi: 10.1080/08870446.2017.1341516] [Medline: 28681612]
21. Lippke S, Ziegelmann JP, Schwarzer R. Behavioral intentions and action plans promote physical exercise: a longitudinal study with orthopedic rehabilitation patients. J Sport Exerc Psychol 2004 Sep;26(3):470-483. [doi: 10.1123/jsep.26.3.470]
22. Zhang CQ, Zhang R, Schwarzer R, Hagger MS. A meta-analysis of the health action process approach. Health Psychol 2019 Jul;38(7):623-637. [doi: 10.1037/hea0000728] [Medline: 30973747]
23. Evers KE, Quintiliani LM. Advances in multiple health behavior change research. Transl Behav Med 2013 Mar;3(1):59-61 [FREE Full text] [doi: 10.1007/s13142-013-0198-z] [Medline: 24073161]

24. Geller K, Lippke S, Nigg CR. Future directions of multiple behavior change research. J Behav Med 2017 Feb;40(1):194-202. [doi: 10.1007/s10865-016-9809-8] [Medline: 27785652]
25. Sweet SN, Fortier MS. Improving physical activity and dietary behaviours with single or multiple health behaviour interventions? A synthesis of meta-analyses and reviews. Int J Environ Res Public Health 2010 Apr;7(4):1720-1743 [FREE Full text] [doi: 10.3390/ijerph7041720] [Medline: 20617056]
26. Lippke S, Nigg CR, Maddock JE. Health-promoting and health-risk behaviors: theory-driven analyses of multiple health behavior change in three international samples. Int J Behav Med 2012 Mar;19(1):1-13 [FREE Full text] [doi: 10.1007/s12529-010-9135-4] [Medline: 21234735]
27. Conner M, Norman P. Health behaviour: current issues and challenges. Psychol Health 2017 Aug;32(8):895-906. [doi: 10.1080/08870446.2017.1336240] [Medline: 28612656]
28. Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE. Effects of eHealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021 Feb 22;23(2):e23786 [FREE Full text] [doi: 10.2196/23786] [Medline: 33616534]
29. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
30. Prestwich A, Webb TL, Conner M. Using theory to develop and test interventions to promote changes in health behaviour: evidence, issues, and recommendations. Curr Opin Psychol 2015 Oct;5:1-5. [doi: 10.1016/j.copsyc.2015.02.011]
31. Lippke S, Ziegelmann JP. Theory-based health behavior change: developing, testing, and applying theories for evidence-based interventions. Appl Psychol 2008 Oct;57(4):698-716. [doi: 10.1111/j.1464-0597.2008.00339.x]
32. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. Ann Behav Med 2018 May 18;52(6):501-512 [FREE Full text] [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
33. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474. [doi: 10.1037/a0032853] [Medline: 23730717]
34. Duan Y, Liang W, Wang Y, Lippke S, Lin Z, Shang B, et al. The effectiveness of sequentially delivered Web-based interventions on promoting physical activity and fruit-vegetable consumption among Chinese college students: mixed methods study. J Med Internet Res 2022 Jan 26;24(1):e30566 [FREE Full text] [doi: 10.2196/30566] [Medline: 35080497]
35. Schwarzer R. Health Action Process Approach (HAPA) as a theoretical framework to understand behavior change. Actualidades Psicología 2016 Dec 05;30(121):119-130. [doi: 10.15517/ap.v30i121.23458]
36. Paech J, Luszczynska A, Lippke S. A rolling stone gathers no moss-the long way from good intentions to physical activity mediated by planning, social support, and self-regulation. Front Psychol 2016 Jul 6;7:1024 [FREE Full text] [doi: 10.3389/fpsyg.2016.01024] [Medline: 27458417]
37. Liang W, Duan YP, Shang BR, Wang YP, Hu C, Lippke S. A web-based lifestyle intervention program for Chinese college students: study protocol and baseline characteristics of a randomized placebo-controlled trial. BMC Public Health 2019 Aug 13;19(1):1097 [FREE Full text] [doi: 10.1186/s12889-019-7438-1] [Medline: 31409381]
38. Duan YP, Wienert J, Hu C, Si GY, Lippke S. Web-based intervention for physical activity and fruit and vegetable intake among Chinese university students: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e106 [FREE Full text] [doi: 10.2196/jmir.7152] [Medline: 28396306]
39. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. BMJ 2015 Sep 04;351:h4672. [doi: 10.1136/bmj.h4672] [Medline: 26341898]
40. Macfarlane DJ, Lee CC, Ho EY, Chan KL, Chan DT. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). J Sci Med Sport 2007 Feb;10(1):45-51. [doi: 10.1016/j.jsams.2006.05.003] [Medline: 16807105]
41. Rafferty AP, Anderson JV, McGee HB, Miller CE. A healthy diet indicator: quantifying compliance with the dietary guidelines using the BRFSS. Prev Med 2002 Jul;35(1):9-15. [doi: 10.1006/pmed.2002.1056] [Medline: 12079436]
42. Lippke S, Fleig L, Wiedemann AU, Schwarzer R. A computerized lifestyle application to promote multiple health behaviors at the workplace: testing its behavioral and psychological effects. J Med Internet Res 2015 Oct 01;17(10):e225 [FREE Full text] [doi: 10.2196/jmir.4486] [Medline: 26429115]
43. Duan YP, Liang W, Guo L, Wienert J, Si GY, Lippke S. Evaluation of a Web-based intervention for multiple health behavior changes in patients with coronary heart disease in home-based rehabilitation: pilot randomized controlled trial. J Med Internet Res 2018 Nov 19;20(11):e12052 [FREE Full text] [doi: 10.2196/12052] [Medline: 30455167]
44. Lippke S, Ziegelmann JP, Schwarzer R, Velicer WF. Validity of stage assessment in the adoption and maintenance of physical activity and fruit and vegetable consumption. Health Psychol 2009 Mar;28(2):183-193 [FREE Full text] [doi: 10.1037/a0012983] [Medline: 19290710]
45. Luszczynska A, Schwarzer R. Planning and self-efficacy in the adoption and maintenance of breast self-examination: a longitudinal study on self-regulatory cognitions. Psychol Health 2003 Jan;18(1):93-108. [doi: 10.1080/0887044021000019358]

46. Reinwand DA, Crutzen R, Storm V, Wienert J, Kuhlmann T, de Vries H, et al. Generating and predicting high quality action plans to facilitate physical activity and fruit and vegetable consumption: results from an experimental arm of a randomised controlled trial. BMC Public Health 2016 Apr 12;16:317 [FREE Full text] [doi: 10.1186/s12889-016-2975-3] [Medline: 27066779]
47. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan 10;363(9403):157-163. [doi: 10.1016/S0140-6736(03)15268-3] [Medline: 14726171]
48. Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken, NJ, USA: John Wiley & Sons; 2009.
49. Liu S, Tu D. On the applications of fisher’s least significant difference (LSD) procedure in three-arm clinical trials with survival endpoints. Drug Information J 2008 Dec 30;42(1):81-91 [FREE Full text] [doi: 10.1177/009286150804200112]
50. Castro-Schilo L, Grimm KJ. Using residualized change versus difference scores for longitudinal research. J Soc Pers Relation 2017 Dec 19;35(1):32-58. [doi: 10.1177/0265407517718387]
51. Schroeder MA. Diagnosing and dealing with multicollinearity. West J Nurs Res 1990 Apr;12(2):175-187. [doi: 10.1177/019394599001200204] [Medline: 2321373]
52. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ, USA: Hillsdale Lawrence Erlbaum; 1969.
53. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
54. Cotter DA, Hermsen JM, Ovadia S, Vanneman R. The glass ceiling effect. Soc Forces 2001;80(2):655-681 [FREE Full text] [doi: 10.1353/sof.2001.0091]
55. Wyse R, Wolfenden L, Bisquera A. Characteristics of the home food environment that mediate immediate and sustained increases in child fruit and vegetable consumption: mediation analysis from the Healthy Habits cluster randomised controlled trial. Int J Behav Nutr Phys Act 2015 Sep 17;12:118 [FREE Full text] [doi: 10.1186/s12966-015-0281-6] [Medline: 26381609]
56. Brug J, Oenema A, Ferreira I. Theory, evidence and Intervention Mapping to improve behavior nutrition and physical activity interventions. Int J Behav Nutr Phys Act 2005 Apr 04;2(1):2 [FREE Full text] [doi: 10.1186/1479-5868-2-2] [Medline: 15807898]
57. Liu HX, Chow BC, Liang W, Hassel H, Huang YW. Measuring a broad spectrum of eHealth skills in the Web 3.0 context using an eHealth Literacy Scale: development and validation study. J Med Internet Res 2021 Sep 23;23(9):e31627 [FREE Full text] [doi: 10.2196/31627] [Medline: 34554098]
58. Nezu AM, Nezu CM. Evidence-Based Outcome Research: A Practical Guide to Conducting Randomized Controlled Trials for Psychosocial Interventions. Oxford, UK: Oxford University Press; 2007.
59. Stok FM, Renner B, Clarys P, Lien N, Lakerveld J, Deliens T. Understanding eating behavior during the transition from adolescence to young adulthood: a literature review and perspective on future research directions. Nutrients 2018 May 24;10(6):667 [FREE Full text] [doi: 10.3390/nu10060667] [Medline: 29794986]
60. Rhodes RE. The evolving understanding of physical activity behavior: a multi-process action control approach. Adv Motiv Sci 2017;4:171-205. [doi: 10.1016/bs.adms.2016.11.001]

## Abbreviations

**CONSORT:** Consolidated Standards of Reporting Trials
**FVC:** fruit and vegetable consumption
**HAPA:** health action process approach
**MHBC:** multiple health behavior change
**PA:** physical activity
**RCT:** randomized controlled trial
**VAS:** visual analog scale

*Edited by G Eysenbach; submitted 25.02.22; peer-reviewed by L Guo, A Videira-Silva, K Uludag, N Maglaveras; comments to author 02.05.22; revised version received 09.06.22; accepted 28.07.22; published 07.09.22.*

*Please cite as:*
*Liang W, Duan Y, Wang Y, Lippke S, Shang B, Lin Z, Wulff H, Baker JS*
*Psychosocial Mediators of Web-Based Interventions for Promoting a Healthy Lifestyle Among Chinese College Students: Secondary Analysis of a Randomized Controlled Trial*
*J Med Internet Res 2022;24(9):e37563*
*URL: https://www.jmir.org/2022/9/e37563*
*doi:10.2196/37563*
*PMID:36069840*

Original Paper

# Internet-Delivered Cognitive Behavioral Therapy for Postsecondary Students: Randomized Factorial Trial for Examining Motivational Interviewing and Booster Lessons

Vanessa Peynenburg[1], MA; Heather Hadjistavropoulos[1], PhD; David Thiessen[1], BSc; Nickolai Titov[2], PhD; Blake Dear[2], PhD

[1]Department of Psychology, University of Regina, Regina, SK, Canada
[2]eCentre Clinic, Department of Psychology, Macquarie University, Sydney, Australia

**Corresponding Author:**
Heather Hadjistavropoulos, PhD
Department of Psychology
University of Regina
3737 Wascana Parkway Drive
Regina, SK, S4S 0A2
Canada
Phone: 1 306 585 5133
Email: hadjista@uregina.ca

## Abstract

**Background:** Internet-delivered cognitive behavioral therapy (ICBT) can improve access to mental health care for students, although high attrition rates are concerning and little is known about long-term outcomes. Motivational interviewing (MI) exercises and booster lessons can improve engagement and outcomes in face-to-face cognitive behavioral therapy.

**Objective:** This study aimed to examine the use of pretreatment MI exercises and booster lessons in ICBT for postsecondary students.

**Methods:** In this factorial trial (factor 1: web-based MI before treatment; factor 2: self-guided booster lesson 1 month after treatment), 308 clients were randomized to 1 of 4 treatment conditions, with 277 (89.9%) clients starting treatment. All clients received a 5-week transdiagnostic ICBT course (the *UniWellbeing* course). Primary outcomes included changes in depression, anxiety, and perceived academic functioning from before treatment to after treatment and at the 1-month and 3-month follow-ups.

**Results:** Overall, 54% (150/277) of students completed treatment and reported large improvements in symptoms of depression and anxiety and small improvements in academic functioning after treatment, which were maintained at the 1-month and 3-month follow-ups. Pretreatment MI did not contribute to better treatment completion or engagement, although small between-group effects favored MI for reductions in depression (Cohen $d$=0.23) and anxiety (Cohen $d$=0.25) after treatment. Only 30.9% (43/139) of students randomized to one of the booster conditions accessed the booster. Overall, no main effects were found for the booster. Subanalyses revealed that clients who accessed the booster had larger decreases in depressive symptoms (Cohen $d$=0.31) at the 3-month follow-up. No interactions were found between MI and the booster.

**Conclusions:** Rather than offering MI before treatment, clients may experience more benefits from MI exercises later in ICBT when motivation wanes. The low uptake of the self-guided booster limited our conclusions regarding its effectiveness. Future research should examine offering a booster for a longer duration after treatment, with therapist support and a longer follow-up period.

**Trial Registration:** ClinicalTrials.gov NCT04264585; https://clinicaltrials.gov/ct2/show/NCT04264585

***(J Med Internet Res 2022;24(9):e40637)*** doi:10.2196/40637

**KEYWORDS**

postsecondary students; transdiagnostic; boosters; motivational interviewing; internet-delivered cognitive behavioral therapy

# Introduction

## Background

An estimated one-third of college students worldwide meet the criteria for mental health disorders in any given year [1]. Depression, generalized anxiety disorder (GAD), and alcohol misuse are the most common mental health concerns among students and are associated with academic difficulties [2], distress [3], and increased risk of suicide [1]. During the COVID-19 pandemic, postsecondary students experienced many additional stressors that may contribute to poor mental health, including uncertainty about the future [4], adoption of new caregiving roles [5], limited opportunities for social contact [6], and strain on personal relationships [7]. Unfortunately, students face numerous attitudinal and structural barriers that prevent them from receiving treatment [8]. It is estimated that only approximately one-third of college students who meet the criteria for a mental health disorder in a given year receive some form of treatment [9].

Internet-delivered cognitive behavioral therapy (ICBT) is an effective alternative to face-to-face cognitive behavioral therapy (CBT) and may reduce the barriers that students face when trying to access treatment. In ICBT, clients receive structured web-based materials (eg, presentation slides, worksheets, and homework activities) based on cognitive behavioral strategies. Therapist-assisted ICBT courses typically involve weekly contact with a therapist either through secure messaging or telephone [10]. In a meta-analysis of internet interventions for university students, Harrer et al [11] found that interventions based on CBT had pooled effect sizes of $g$=0.29 for depression, $g$=0.39 for anxiety, and $g$=0.33 for stress. Overall, these effect sizes are smaller than what is reported in studies of ICBT within the general population [10], which suggests that strategies for improving the outcomes of ICBT among college and university students should be explored.

Another issue of ICBT for university students is high attrition rates [12]. In a study on predictors of outcomes in ICBT, it was found that younger clients were more likely to drop out of ICBT than older clients [13]. A possible explanation for high attrition rates among students is that they experience low motivation or ambivalence to change [14]. In face-to-face settings, low motivation and ambivalence to change can be addressed using motivational interviewing (MI) [15], and there is evidence for improved treatment acceptability, adherence [15], and treatment outcomes [16] in studies that combine MI with CBT [15]. Even single-session interventions involving MI have been found to produce larger effects than interventions that do not include MI in studies on reducing heavy drinking in college students [17]. Furthermore, brief MI interventions as short as 15 minutes can contribute to behavior changes [18]. Including a series of MI-based exercises at the beginning of ICBT may be beneficial to improving client outcomes.

The literature on adding MI to ICBT is limited, with only 2 studies exploring MI before clients initiate ICBT [19,20]. In 1 study, 108 clients with social anxiety were randomized to self-guided ICBT with or without a web-based MI lesson before treatment [20]. The MI lesson comprised reflective questions that encouraged clients to consider the short- and long-term costs and benefits of ICBT, the barriers they might encounter during ICBT, and the strengths they could draw on to overcome those barriers. Clients assigned to the MI lesson group were more likely to complete the full course of treatment than those who received ICBT alone (75% vs 56% completion), although no differences in outcomes were found between the groups. In another trial, 480 clients were randomized to either receive or not receive a pretreatment MI lesson before accessing ICBT [19]. The MI lesson comprised 3 videos and 5 web-based exercises with open-ended, reflective questions. Clients who received the MI lesson engaged in more change talk in their email exchanges with therapists and were enrolled in the course for more days than those who did not receive MI. However, there was no evidence that pretreatment MI contributed to better treatment completion rates or greater reductions in symptoms of depression and anxiety. The authors identified that motivation was high among clients before treatment and therefore speculated that benefits may not have been observed as clients did not generally experience low motivation.

To date, no studies have examined the inclusion of pretreatment MI in ICBT for postsecondary students, a group known to be at risk of lower engagement and poorer outcomes. Given the promising findings of including MI in face-to-face CBT [16] and preliminary findings for improved engagement in ICBT [19,20], it is worthwhile to explore the use of pretreatment MI to address high attrition rates among students.

Booster sessions represent another strategy with the potential to improve ICBT outcomes. In face-to-face CBT, booster sessions are used to remind clients of strategies learned during treatment and offer clients the opportunity to problem solve any barriers faced since completing treatment [21]. Booster sessions appear to help clients maintain treatment gains for a longer period of time than with treatment alone [21,22]. In the case of ICBT, there is some evidence showing that booster sessions after ICBT contribute to greater levels of overall functioning and delay the onset of relapse in clients with obsessive-compulsive disorder [23]. However, in another recent study of booster lessons with ICBT, the benefits of a booster session could not be adequately explored as uptake of the therapist-assisted booster lesson was low [24]. Overall, given the need to improve the engagement and outcomes of ICBT among postsecondary students, the benefits of booster lessons after ICBT are worthy of exploration.

## Objective

The purpose of this study was to examine the effects of including pretreatment MI and a self-guided booster offered 1 month after transdiagnostic ICBT for postsecondary students. In particular, we were interested in whether including pretreatment MI would affect treatment completion, treatment engagement, and outcomes compared with ICBT without pretreatment MI. We hypothesized that clients assigned to either of the conditions with pretreatment MI would be more likely to complete treatment and be more engaged during treatment (ie, greater number of log-ins and more messages sent to their therapist) than clients who did not receive the pretreatment MI exercises (ie, only received the standard ICBT course or the ICBT course with a self-guided booster lesson). The study of the benefits of pretreatment MI for symptom improvement was considered exploratory in this specific population, given the null findings from both previous ICBT studies regarding the benefits of pretreatment MI [19,20]. We were also interested in examining students' use of a self-guided booster lesson and how the use of the booster would affect outcomes at the 3-month follow-up. It was hypothesized that clients assigned to the booster would have larger improvements than those not assigned to the booster.

## Methods

### Study Design

This study used a 2×2 factorial design (factor 1: pretreatment web-based MI; factor 2: self-guided booster lesson 1 month after treatment), and was registered as a clinical trial (ClinicalTrials.gov NCT04264585).

### Ethics Approval

The study was reviewed and received ethics approval from the University of Regina Research Ethics Board (REB: 2019-205).

### Setting

The study was conducted at a routine care ICBT clinic (the Online Therapy Unit), which offers ICBT free of charge to residents of Saskatchewan. Examining client outcomes in a routine care setting is important as these clients typically present with greater levels of comorbidity and more diversity than in early phase randomized controlled trials [25]. To identify small between-group effect sizes with a power of 80% and an α of .10, a minimum of 277 participants was required. An α of .10 has been suggested for optimization studies [26] and has been used in previous trials of ICBT to identify active factors [27]. Once a factor has been confirmed as important for improving outcomes in the confirmatory phase, a traditional α value of .05 can be used [26].

### Clients

Prospective clients could self-refer to the course using the Online Therapy Unit's website. Clients found out about the *UniWellbeing* course through various sources, as described in the *Results* section. To be eligible, clients had to (1) be younger than 18 years, (2) self-report at least mild symptoms of depression or anxiety (ie, score ≥5) on the 9-item Patient Health Questionnaire (PHQ-9) [28] or 7-item Generalized Anxiety Disorder (GAD-7) questionnaire [29], (3) be enrolled at a postsecondary institution in the province of Saskatchewan, (4) have reliable access to a computer and the internet and feel comfortable using them, and (5) remain in Saskatchewan during the treatment period. Prospective clients were ineligible if they (1) had unmanaged psychosis, mania, or alcohol or drug problems; (2) were hospitalized in the previous year for mental health; (3) were at high suicide risk; or (4) did not meet the eligibility criteria listed previously.

### Intervention

All clients were offered the *UniWellbeing* course, which is a 5-week transdiagnostic, therapist-assisted ICBT course designed for postsecondary students. The course was developed at Macquarie University in Australia and has undergone several revisions to increase treatment completion [30,31]. In its current form, the *UniWellbeing* course comprises 4 lessons that are released over 5 weeks. Lessons are presented in a slideshow format and include case stories, downloadable lesson summaries, and homework exercises. The 4 lessons include information about the following: symptom identification and the CBT model, thought monitoring and thought challenging, symptoms of over- and underarousal and how they can be managed using breathing exercises and pleasant activity scheduling, avoidance or safety behaviors and graded exposure, and relapse prevention. Throughout the course, clients can access additional resources (ie, assertive communication, communication skills, emergency contact information, grief, managing beliefs, mental skills, and problem-solving), as well as 2 additional case stories that were created for this trial in response to therapist feedback (ie, COVID-19 and mature student case story). Clients receive automated messages to alert them about lessons as they become available and to remind them of how far along they should be in the course.

### Randomization and Factors

Clients who were accepted into the trial were randomized using REDCap (Research Electronic Data Capture; Vanderbilt University) to one of four treatment conditions (using block randomization) at the end of their telephone screen: standard *UniWellbeing*, MI+*UniWellbeing*, *UniWellbeing*+booster, or MI+*UniWellbeing*+booster.

#### *MI Condition*

Clients who were randomized to 1 of the 2 conditions that included MI completed a series of 5 web-based exercises (the *Planning for Change* lesson) before starting the main *UniWellbeing* course. The MI exercises were adapted from previous research [19,32] using previous client feedback, as well as input from 2 patient partners with lived experiences of anxiety or depression. Revisions included removing redundancy and improving the clarity of the exercises (eg, by including examples of strengths). Previous clients suggested that the MI exercises were too long, which could have negatively affected client motivation; therefore, the number of questions in each exercise was reduced.

In the trial by Soucy et al [19], clients completed the MI exercises 1 week before starting treatment, and the exercises were delivered on a separate platform from the main treatment, which prevented therapists from viewing client responses. In this trial, a decision was made to include the MI exercises directly on the treatment platform. All questions within each MI exercise were required for clients to access lesson 1 of the *UniWellbeing* course.

The exercises were based on common MI principles (ie, value clarification [33,34], importance ruler [33], looking back [33,34], confidence ruler [33], and looking forward [33]). Clients completed a series of open-ended questions as part of the exercises and received automated feedback to provide encouragement (eg, "You can do this. Your courage to reach out for help demonstrates that you have already started taking the steps to manage your anxiety and/or depression")

### *Booster Lesson*

Clients who were randomized to the booster condition were offered a self-guided booster lesson 1 month after completing the *UniWellbeing* course. The booster lesson comprised a summary of the key skills from the course, a section on maintaining motivation, and information about structured problem-solving and how it could be used to manage lapses. It was designed to reflect common content in face-to-face booster sessions [21]. The booster was presented in the same slideshow format as the core 4 lessons and included a printable summary with worksheets.

### Therapist Support

Each client was assigned to a therapist who provided weekly support during the 5-week course. Most of the therapist support was offered through personalized messages sent on the secure treatment portal on the same day each week. Therapists were instructed to spend 15 to 20 minutes per client each week. Telephone contact could be initiated if the client experienced a significant increase in symptoms of depression or anxiety (≥5 points on the PHQ-9 or GAD-7), if the client's responses on the PHQ-9 or messages to their therapist suggested increased suicide risk, or if the client had not accessed the website for a week to encourage the client to continue working on the lessons. A total of 6 therapists provided support in the trial (n=2, 33% with backgrounds in psychology, and n=4, 67% in social work). Of the 6 therapists, 5 (83%) were registered with their respective regulatory colleges, and 1 (17%) was a supervised PhD student in clinical psychology. All 6 therapists received ICBT training, regular supervision, and auditing of their messages [35,36]. Therapists were encouraged to include the following in each message to the clients: express warmth and concern, provide feedback on weekly symptom questionnaires, highlight content from lessons, address client questions, reinforce skill acquisition and progress, manage risks, and remind clients of upcoming lesson content and their next check-in.

## Outcomes

### *Primary Outcome Measures*

Primary outcome measures were administered before treatment, after treatment, and at the 1-month and 3-month follow-ups. Clients also completed these measures on a weekly basis during treatment as a way for therapists to monitor their symptoms.

#### 9-Item Patient Health Questionnaire

The PHQ-9 [28] comprises 9 self-report items to assess depressive symptoms over the past 2 weeks. Total scores range from 0 to 27, with a score of ≥10 often used as an indicator of clinically significant symptoms [28]. The PHQ-9 had good internal consistency (α=.83 to α=.86) in this study.

#### 7-Item Generalized Anxiety Disorder

The GAD-7 [29] comprises 7 self-report items to assess symptoms of GAD over the past 2 weeks. Total scores range from 0 to 21, and a score of ≥10 has been used to identify individuals with clinical levels of generalized anxiety [29]. The GAD-7 had good to excellent internal consistency in this study (α=.85 to α=.93).

#### Perceptions of Academic Functioning

The Perceptions of Academic Functioning (PAF) was created for this study and comprises 3 items related to academic functioning over the past week. Using a scale ranging from 0 to 10, clients were asked how well they felt they were able to attend their classes or lectures, complete academic tasks (eg, assignments, papers, laboratory classes, and readings), and absorb information from readings or lectures. The PAF had good internal consistency (α=.81 to α=.93), and items were summed to create a total score, with higher scores indicating better perceived functioning.

### *Secondary Outcome Measures*

Secondary outcome measures were administered at various time points, as described in the following sections:

#### Sheehan Disability Scale

The Sheehan Disability Scale (SDS) [37] comprises 3 self-report items assessing mental health–related disability in the areas of work or school, social, and family life. Total scores range from 0 to 30, with higher scores indicating greater levels of impairment. The SDS was administered before treatment, after treatment, and at the 1-month and 3-month follow-ups and had acceptable to good internal consistency (α=.71 to α=.88).

#### Alcohol Use Disorder Identification Test

The Alcohol Use Disorder Identification Test (AUDIT) [38] is a 10-item self-report measure that includes questions on alcohol consumption. Total scores range between 0 and 40, with a score >20 warranting monitoring and brief intervention [39]. In this trial, the AUDIT was administered before and after treatment and had acceptable to good internal consistency (α=.77 to α=.81).

#### Drug Use Disorder Identification Test

The Drug Use Disorder Identification Test (DUDIT) [40] comprises 11 self-report items related to drug use. Total scores range between 0 and 44, and a score of ≥25 suggests the presence of drug dependence [40]. The DUDIT was administered before treatment and after treatment and had good internal consistency (α=.86 at both time points).

#### The 3-Item Change Questionnaire

The 3-Item Change Questionnaire (CQ-3) [41] comprises 3 items that focus on clients' perceptions of their ability to make changes, the importance of making changes, and the level of commitment to making changes to their symptoms. All clients completed the CQ-3 before treatment, and clients assigned to MI also completed the CQ-3 after the MI exercises. The CQ-3 had poor internal consistency (α=.55 to α=.66).

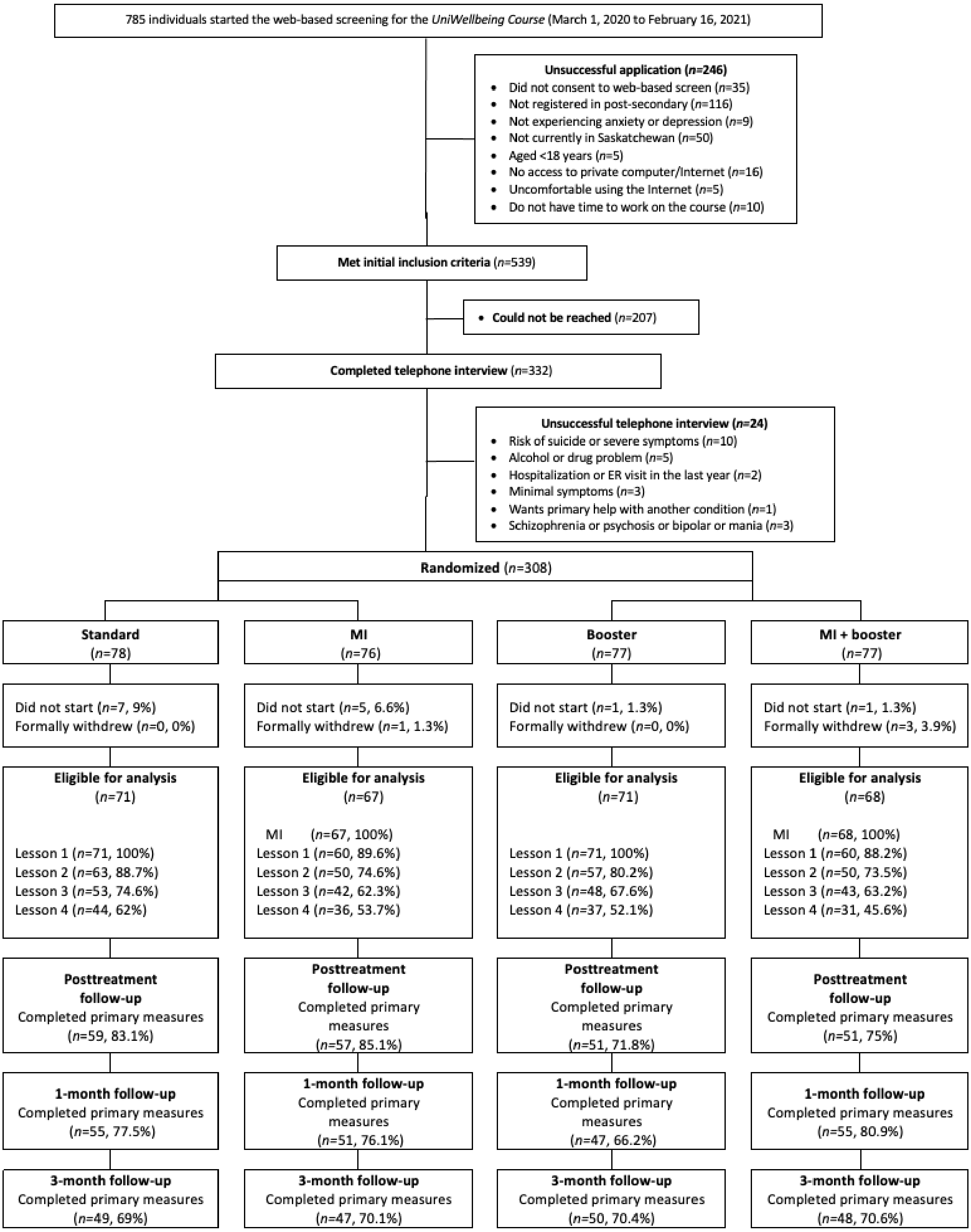
#### Treatment Satisfaction Questionnaire

The Treatment Satisfaction Questionnaire [42] comprises 6 items to assess ICBT treatment satisfaction. Clients complete questions about whether they would refer the course to a friend, whether they thought the course was worth their time (*yes/no*), how satisfied they were with the overall treatment and quality of the course (1=*very dissatisfied* to 5=*very satisfied*), and the extent to which the course affected their confidence in managing their symptoms and motivation to seek help in the future (1=*greatly reduced* to 5=*greatly increased*).

## Analyses

Descriptive statistics were used to summarize client characteristics at intake, including demographics and scores on each of the primary and secondary outcome measures. Modified intention-to-treat (ITT) analyses [43] were used to generate replacement scores for missing data, which included only those clients who started treatment (Figure 1 shows the amount of missing data at each time point). A modified ITT was chosen to maintain consistency with the published trial of the *UniWellbeing* course [31], which allowed us to make more direct comparisons in effect sizes between studies, as opposed to a true ITT. In addition, true ITT analyses may be too conservative when examining effect sizes, as including clients who do not start treatment indicates little about the efficacy of the treatment [43]. A model was created that included 50 imputations and controlled for other values of the outcome measure, treatment condition, course completion, ethnicity, location, use of the booster, and interactions between pretreatment measures and the treatment condition [44-46]. Generalized estimation equations were used to examine changes in primary and secondary outcome measures across time points, consistent with past ICBT research [47,48]. Generalized estimation equation models used an exchangeable working correlation to account for within-subject variance. For the PHQ-9, GAD-7, PAF, and SDS, a gamma distribution model with a log link function was used, which can model improvements as proportional to pretreatment symptom severity and accommodate skewed response distributions [46]. For the AUDIT and DUDIT, the number of *0* value responses caused numerical issues; therefore, a more traditional Gaussian distribution with an identity link was used instead. At each of the 3 time points, pairwise comparisons were used to examine within-group and between-group effect sizes. Reliable improvement and reliable recovery were used as indicators of clinically significant change. Consistent with previous research, a reduction of 6 points on the PHQ-9 and 4 points on the GAD-7 was used as an indicator of reliable improvement [47,48]. Conversely, an increase of 6 points on the PHQ-9 and 4 points on the GAD-7 was an indicator of deterioration [47,48]. To meet the criteria for reliable recovery, clients had to score ≥10 on the PHQ-9 or GAD-7 before treatment, score <10 at the follow-up time point, and report at least a 6-point or 4-point reduction, respectively [47,48]. Finally, a series of ANOVAs and chi-square analyses were conducted to examine group differences in treatment completion, client engagement, and treatment satisfaction.

**Figure 1.** Patient flow from screening to the 3-month follow-up. ER: emergency room; MI: motivational interviewing.



## Results

### Participant Flow and Characteristics

In this trial, 308 clients were randomized (Figure 1). To be considered eligible for analysis (277/308, 89.9%), clients had to consent to treatment and either start the MI exercises (MI conditions) or start lesson 1 (conditions that did not receive MI). The mean age of the sample was 23.73 (SD 5.95) years, and most of the sample were women (224/277, 80.9%) and White (201/277, 72.6%), had full-time student status (235/277, 85.5%), were living with someone (224/277, 80.9%), and were living within 1 of the 2 major cities in the province of Saskatchewan (191/277, 69%). Approximately half of the students (135/277, 48.7%) indicated that they were employed while completing their studies.

Before treatment, the mean scores on the PHQ-9 and GAD-7 were 14.49 (SD 5.77) and 13.71 (SD 4.76), respectively. Over two-thirds of clients scored above the clinical cutoff (≥10) on both the PHQ-9 and GAD-7, and only a small subset of clients did not score within the clinical range on either the PHQ-9 or the GAD-7 (34/277, 12.3%). The mean score on the PAF was 17.36 (SD 6.46). In terms of mental health history and service use, 37.9% (105/277) reported having a mental health diagnosis, 27.4% (76/277) reported having some form of mental health support at screening, and 36.5% (101/277) reported having taken psychotropic medication for anxiety or depression in the previous 3 months. The mean score for the CQ-3 was 24.05 (SD 4.22) before treatment, suggesting relatively high levels of motivation. Table 1 shows a full overview of the clients' pretreatment characteristics.

**Table 1.** Pretreatment demographic and clinical characteristics by group (N=277).

| Variable | All groups | Standard (n=71) | MI[a] (n=67) | Booster (n=71) | MI+booster (n=68) |
|---|---|---|---|---|---|
| Age (years), mean (SD; range) | 23.73 (5.95; 17-46) | 23.57 (6.11; 18-44) | 24.66 (6.58; 18-46) | 23.35 (5.86; 17-43) | 23.44 (5.28; 18-39) |
| **Gender, n (%)** | | | | | |
| Man | 45 (16.2) | 10 (14.1) | 6 (9.1) | 12 (18.2) | 17 (25) |
| Woman | 224 (80.9) | 57 (80.3) | 58 (86.6) | 58 (81.7) | 51 (75) |
| Nonbinary | 5 (1.8) | 2 (2.8) | 2 (3) | 1 (1.5) | 0 (0) |
| Prefer not to disclose or not listed | 3 (1.1) | 2 (2.8) | 1 (1.5) | 0 (0) | 0 (0) |
| **Marital status, n (%)** | | | | | |
| Single or never married | 113 (41.5) | 31 (43.7) | 22 (33.3) | 31 (47) | 29 (42.6) |
| Dating | 107 (38.6) | 26 (36.6) | 24 (35.8) | 29 (40.8) | 28 (41.2) |
| Married or common law | 31 (11.4) | 8 (11.3) | 13 (19.7) | 5 (7.6) | 5 (7.4) |
| Living with partner | 18 (6.4) | 4 (5.6) | 3 (4.5) | 5 (7.6) | 6 (8.8) |
| Separated or divorced or widowed | 8 (3) | 2 (2.8) | 5 (7.6) | 1 (1.5) | 0 (0) |
| Living with someone | 224 (80.9) | 55 (77.5) | 56 (83.6) | 57 (80.3) | 56 (82.4) |
| **Postsecondary institution, n (%)** | | | | | |
| University of Regina | 186 (67.1) | 47 (66.2) | 46 (70.1) | 47 (66.2) | 45 (66.2) |
| University of Saskatchewan | 59 (21.3) | 14 19.7 () | 11 (16.4) | 18 (25.4) | 16 (23.5) |
| Saskatchewan Polytechnic | 10 (3.6) | 1 (1.4) | 6 (9.1) | 1 (1.4) | 2 (2.9) |
| Other | 22 (8.1) | 9 (12.7) | 3 (4.5) | 5 (7) | 5 (7.7) |
| **Student status, n (%)** | | | | | |
| Full-time student | 235 (85.5) | 61 (87.1) | 53 (79.1) | 62 (87.3) | 59 (88.1) |
| Part-time student | 40 (14.5) | 9 (12.9) | 14 (20.9) | 8 (12.7) | 8 (11.9) |
| **Year of studies, n (%)** | | | | | |
| First-year undergraduate | 67 (24.6) | 19 (27.1) | 14 (21.2) | 21 (30.4) | 13 (19.4) |
| Second-year undergraduate | 57 (21) | 13 (18.6) | 17 (25.8) | 8 (11.6) | 19 (28.4) |
| Third-year undergraduate | 62 (22.8) | 18 (25.7) | 17 (25.8) | 16 (23.2) | 11 (16.4) |
| Fourth-year undergraduate | 44 (16.2) | 10 (14.3) | 7 (10.6) | 15 (21.7) | 12 (17.9) |
| Fifth or higher year undergraduate | 18 (6.6) | 3 (4.3) | 7 (10.6) | 4 (5.8) | 4 (6) |
| Graduate or professional student | 26 (8.8) | 7 (10) | 4 (6.1) | 5 (7.2) | 8 (12) |
| International student, n (%) | 10 (3.6) | 3 (4.3) | 2 (3) | 1 (1.4) | 4 (5.9) |
| English not the first language, n (%) | 24 (8.7) | 2 (2.8) | 5 (7.5) | 6 (8.6) | 11 (16.2) |
| **Employment status, n (%)** | | | | | |
| Paid work | 135 (48.7) | 39 (54.9) | 32 (47.8) | 30 (44.1) | 30 (44.1) |
| Unemployed | 142 (51.3) | 32 (45.1) | 35 (52.2) | 38 (55.9) | 38 (55.9) |
| **Ethnicity, n (%)** | | | | | |
| White | 201 (72.6) | 53 (74.6) | 49 (73.1) | 54 (76.1) | 45 (66.2) |
| Indigenous | 20 (7.2) | 6 (8.5) | 8 (11.9) | 3 (4.2) | 3 (4.4) |
| Asian | 26 (9.4) | 3 (4.2) | 4 (6) | 8 (11.2) | 11 (16.2) |
| Other | 30 (10.8) | 9 (12.7) | 6 (9) | 6 (8.5) | 9 (13.2) |
| **Location[b], n (%)** | | | | | |
| Large city (>200,000) | 191 (69) | 49 (69) | 44 (65.7) | 52 (73.2) | 46 (67.6) |
| Small to medium city | 23 (8.3) | 8 (11.3) | 6 (9) | 7 (9.9) | 2 (2.9) |

| Variable | All groups | Standard (n=71) | MI[a] (n=67) | Booster (n=71) | MI+booster (n=68) |
|---|---|---|---|---|---|
| Small rural location (<10,000) | 63 (22.7) | 14 (19.7) | 17 (25.4) | 12 (16.9) | 20 (29.4) |
| **Referral source, n (%)** | | | | | |
| Physician or medical professional | 88 (31.9) | 19 (27.1) | 23 (34.3) | 22 (31) | 24 (35.3) |
| Web-based source (eg, website or email) | 70 (25.4) | 19 (27.1) | 17 (25.4) | 20 (28.2) | 14 (20.6) |
| Counseling services | 37 (13.4) | 10 (14.3) | 11 (16.4) | 6 (8.5) | 10 (14.7) |
| Friend or family member or employer | 35 (12.7) | 6 (8.6) | 10 (14.9) | 12 (16.9) | 7 (10.3) |
| Presentation | 12 (4.3) | 5 (7.1) | 2 (2.8) | 2 (2.8) | 3 (4.4) |
| Printed poster or media | 4 (1.4) | 4 (5.7) | 0 (0) | 0 (0) | 1 (1.5) |
| Other | 30 (10.9) | 7 (10) | 4 (6) | 9 (12.7) | 8 (11.8) |
| **Mental health characteristics, n (%)** | | | | | |
| Lifetime mental health service use | 158 (57) | 36 (50.7) | 47 (70.1) | 38 (53.5) | 37 (54.4) |
| Lifetime hospitalization for mental health | 22 (7.9) | 3 (4.2) | 9 (13.4) | 7 (9.9) | 3 (4.4) |
| Mental health diagnosis | 105 (37.9) | 27 (38) | 32 (47.8) | 27 (38) | 19 (27.7) |
| Taking psychotropic medication in the past 3 months | 101 (36.5) | 20 (28.2) | 29 (43.3) | 23 (32.4) | 29 (42.6) |
| Current mental health service use | 76 (27.4) | 19 (26.8) | 23 (34.3) | 12 (16.9) | 22 (32.4) |
| Pretreatment GAD-7[c] ≥10 | 216 (78) | 58 (81.7) | 49 (73.1) | 55 (77.5) | 54 (79.4) |
| Pretreatment PHQ-9[d] ≥10 | 213 (76.9) | 53 (74.6) | 52 (77.6) | 55 (77.5) | 53 (77.9) |
| Pretreatment GAD-7 ≥10 and PHQ-9 ≥10 | 186 (67.1) | 48 (67.6) | 44 (65.7) | 47 (66.2) | 47 (69.1) |
| No clinical scores | 34 (12.3) | 8 (11.3) | 10 (14.9) | 8 (11.3) | 8 (11.8) |
| **Treatment expectations , mean (SD)** | | | | | |
| CEQ[e] | 21.16 (4.19) | 20.63 (3.98) | 21.91 (4.09) | 20.90 (4.61) | 21.34 (4.09) |
| CQ-3[f] | 24.05 (4.22) | 24.29 (3.64) | 23.39 (4.51) | 24.20 (4.46) | 24.35 (4.41) |

[a] MI: motivational interviewing.

[b] Location is based on where the client was residing at intake.

[c] GAD-7: 7-item Generalized Anxiety Disorder.

[d] PHQ-9: 9-item Patient Health Questionnaire.

[e] CEQ: Credibility and Expectancy Questionnaire.

[f] CQ-3: 3-item Change Questionnaire.

## Primary Outcome Measures

The estimated marginal means, percentage reductions, and effect sizes for each of the primary outcome measures are presented in Table 2, separated by factor (MI vs no MI; booster vs no booster). The same information can be found in Multimedia Appendix 1 for each of the 4 treatment conditions (*UniWellbeing*, MI+*UniWellbeing*, *UniWellbeing*+booster, or MI+*UniWellbeing*+booster). Regardless of factors, there were statistically significant time effects for the PHQ-9, GAD-7, and PAF from before treatment to all subsequent measurement periods (after treatment and at the 1-month and 3-month follow-ups). Clients experienced large reductions on the PHQ-9 (Cohen *d*=1.28-1.48) and GAD-7 (Cohen *d*=1.46-1.72) from before treatment to after treatment, with improvements maintained at the 1-month (PHQ-9: Cohen *d*=1.27-1.37; GAD-7: Cohen *d*=1.29-1.51) and 3-month follow-ups (PHQ-9: Cohen *d*=1.22-1.31; GAD-7: Cohen *d*=1.19-1.31).

**Table 2.** Estimated marginal means, 95% CIs, percentage changes, and effect sizes (Cohen *dd*) for primary and secondary outcomes by factor (MI[a] and booster) using pooled imputations.

| Outcomes | Estimated marginal means, (SD) | | | | Changes from pretreatment (%), 95% CI | | | Within-group effect sizes from pretreatment, 95% CI | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pretreatment | Posttreatment | 1 month | 3 months | To posttreatment | To 1 month | To 3 months | To posttreatment | To 1 month | To 3 months |
| **Primary outcomes** | | | | | | | | | | |
| **PHQ-9[b]** | | | | | | | | | | |
| MI | 14.24 (5.56) | 6.90 (4.24) | 7.04 (4.92) | 7.41 (4.78) | 52 (46 to 57) | 51 (44 to 57) | 48 (42 to 54) | 1.48 (1.21 to 1.75) | 1.37 (1.10 to 1.63) | 1.31 (1.05 to 1.58) |
| No MI | 14.72 (5.97) | 7.91 (4.54) | 7.73 (5.01) | 7.79 (5.31) | 46 (41 to 52) | 48 (41 to 54) | 47 (41 to 53) | 1.28 (1.02 to 1.54) | 1.27 (1.01 to 1.52) | 1.22 (0.97 to 1.48) |
| Booster | 14.16 (5.16) | 7.77 (4.41) | 7.60 (5.00) | 7.59 (4.78) | 45 (40 to 51) | 46 (40 to 53) | 46 (40 to 53) | 1.33 (1.07 to 1.59) | 1.29 (1.03 to 1.55) | 1.32 (1.06 to 1.58) |
| No booster | 14.80 (6.33) | 7.03 (4.41) | 7.16 (4.95) | 7.60 (5.33) | 53 (47 to 58) | 52 (45 to 58) | 49 (42 to 55) | 1.42 (1.16 to 1.69) | 1.34 (1.08 to 1.60) | 1.23 (0.97 to 1.48) |
| **GAD-7[c]** | | | | | | | | | | |
| MI | 13.51 (4.71) | 6.10 (3.89) | 6.78 (4.75) | 7.17 (5.42) | 55 (50 to 60) | 50 (43 to 56) | 47 (40 to 54) | 1.71 (1.43 to 1.99) | 1.42 (1.15 to 1.69) | 1.25 (0.98 to 1.51) |
| No MI | 13.88 (4.82) | 7.14 (4.28) | 7.11 (4.94) | 7.51 (5.33) | 49 (43 to 54) | 49 (42 to 55) | 46 (39 to 53) | 1.47 (1.21 to 1.74) | 1.38 (1.12 to 1.64) | 1.25 (1.00 to 1.50) |
| Booster | 13.66 (4.75) | 7.12 (4.19) | 7.34 (5.00) | 7.59 (5.42) | 48 (42 to 54) | 46 (40 to 53) | 44 (37 to 52) | 1.46 (1.19 to 1.72) | 1.29 (1.04 to 1.55) | 1.19 (0.93 to 1.44) |
| No booster | 13.73 (4.80) | 6.12 (4.00) | 6.57 (4.68) | 7.10 (5.32) | 55 (50 to 60) | 52 (46 to 58) | 48 (41 to 55) | 1.72 (1.44 to 1.99) | 1.51 (1.24 to 1.77) | 1.31 (1.05 to 1.57) |
| **PAF[d]** | | | | | | | | | | |
| MI | 17.47 (6.27) | 19.63 (7.30) | 20.14 (6.73) | 19.76 (7.62) | 17 (7 to 28) | 21 (11 to 32) | 18 (7 to 30) | 0.32 (0.08, 0.56) | 0.41 (0.17 to 0.65) | 0.33 (0.09 to 0.57) |
| No MI | 17.24 (6.67) | 19.24 (7.33) | 20.38 (6.09) | 19.12 (7.44) | 16 (5 to 27) | 25 (15 to 34) | 15 (4 to 26) | 0.28 (0.05 to 0.52) | 0.49 (0.25 to 0.73) | 0.27 (0.03 to 0.50) |
| Booster | 17.29 (6.75) | 19.81 (6.94) | 20.52 (6.37) | 19.41 (7.36) | 20 (9 to 31) | 25 (16 to 35) | 17 (6 to 28) | 0.37 (0.13 to 0.61) | 0.49 (0.25 to 0.73) | 0.30 (0.06 to 0.54) |
| No booster | 17.43 (6.19) | 19.05 (7.65) | 19.99 (6.45) | 19.48 (7.71) | 13 (2 to 24) | 20 (10 to 31) | 16 (5 to 28) | 0.23 (–0.01 to 0.47) | 0.40 (0.17 to 0.64) | 0.29 (0.06 to 0.53) |
| **Secondary outcomes** | | | | | | | | | | |
| **SDS[e]** | | | | | | | | | | |
| MI | 19.28 (6.70) | 13.19 (6.79) | 11.99 (7.27) | 12.37 (7.46) | 32 (25 to 38) | 38 (31 to 45) | 36 (29 to 43) | 0.90 (0.65 to 1.15) | 1.04 (0.79 to 1.30) | 0.97 (0.72 to 1.22) |
| No MI | 20.06 (5.75) | 15.63 (7.00) | 12.54 (7.09) | 12.24 (7.38) | 22 (16 to 28) | 37 (31 to 44) | 39 (32 to 46) | 0.69 (0.45 to 0.93) | 1.16 (0.91 to 1.41) | 1.18 (0.93 to 1.43) |
| Booster | 19.58 (6.22) | 14.52 (6.53) | 12.27 (7.25) | 12.18 (7.13) | 26 (19 to 32) | 37 (30 to 44) | 38 (31 to 45) | 0.79 (0.55 to 1.03) | 1.08 (0.83 to 1.33) | 1.10 (0.85 to 1.35) |
| No booster | 19.76 (6.26) | 14.20 (7.45) | 12.25 (7.12) | 12.43 (7.70) | 28 (21 to 35) | 38 (31 to 45) | 37 (30 to 44) | 0.81 (0.56 to 1.05) | 1.12 (0.86 to 1.37) | 1.04 (0.79 to 1.29) |
| **AUDIT[f]** | | | | | | | | | | |
| MI | 4.59 (4.86) | 3.08 (3.95) | —[g] | — | 33 (18 to 48) | — | — | 0.34 (0.10 to 0.58) | — | — |
| No MI | 4.29 (4.06) | 3.52 (3.90) | — | — | 18 (2 to 34) | — | — | 0.19 (–0.04 to 0.43) | — | — |
| Booster | 4.30 (4.28) | 3.28 (3.86) | — | — | 24 (7 to 40) | — | — | 0.25 (0.01 to 0.48) | — | — |

| Outcomes | Estimated marginal means, (SD) | | | | Changes from pretreatment (%), 95% CI | | | Within-group effect sizes from pretreatment, 95% CI | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pretreatment | Posttreatment | 1 month | 3 months | To posttreatment | To 1 month | To 3 months | To posttreatment | To 1 month | To 3 months |
| No booster | 4.58 (4.65) | 3.32 (4.00) | — | — | 28 (12 to 43) | — | — | 0.29 (0.05 to 0.53) | — | — |
| **DUDIT**[h] | | | | | | | | | | |
| MI | 2.32 (5.10) | 1.95 (4.72) | — | — | 16 (20 to 52) | — | — | 0.07 (–0.16 to 0.31) | — | — |
| No MI | 2.90 (6.04) | 2.36 (5.21) | — | — | 19 (–12 to 49) | — | — | 0.10 (–0.14 to 0.33) | — | — |
| Booster | 2.52 (5.09) | 1.66 (3.65) | — | — | 34 (8 to 60) | — | — | 0.19 (–0.04 to 0.43) | — | — |
| No booster | 2.70 (6.09) | 2.66 (6.00) | — | — | 2 (–36 to 40) | — | — | 0.01 (–0.23 to 0.24) | — | — |

[a] MI: motivational interviewing.

[b] PHQ-9: 9-item Patient Health Questionnaire.

[c] GAD-7: 7-item Generalized Anxiety Disorder.

[d] PAF: Perceptions of Academic Functioning.

[e] SDS: Sheehan Disability Scale.

[f] AUDIT: Alcohol Use Disorder Identification Test.

[g] The AUDIT and DUDIT were only administered before treatment and after treatment; thus, data are not available for the percentage change and effect sizes at the 1-month and 3-month follow-ups.

[h] DUDIT: Drug Use Disorder Identification Test.

A main effect was found for MI on the PHQ-9 (between-group Cohen *d*=0.23, 95% CI −0.01 to 0.47; *P*=.06) and GAD-7 (between-group Cohen *d*=0.25, 95% CI 0.02-0.49; *P*=.04) after treatment, whereby clients who were randomized to the MI condition had larger reductions in both measures from pretreatment to posttreatment time points. Between-group differences were no longer significant at the 1-month or 3-month follow-ups for the PHQ-9 (*P*=.25-.52) or GAD-7 (*P*=.57-.60). No significant between-group differences were found for the PAF at any of the 3 time points (*P*=.48-.75).

No main effects were found for those assigned to the booster versus those who were not assigned to any of the primary (*P*=.45-.99) or secondary measures (*P*=.03) at the 3-month follow-up. Owing to low booster use, subanalyses compared clients who did and did not access the booster and found the main effects in favor of accessing the booster on the PHQ-9 (*P*=.09) and PAF (*P*=.02). Clients who accessed the booster had larger improvements in depression (between-group Cohen *d*=0.31, 95% CI −0.05 to 0.67) and perceived academic functioning (between-group Cohen *d*=0.42, 95% CI 0.06-0.78) at the 3-month follow-up. Of note, clients who accessed the booster had higher perceived academic functioning before treatment (between-group Cohen *d*=0.33, 95% CI −0.03 to 0.69; *P*=.08), after treatment (between-group Cohen *d*=0.47, 95% CI 0.11-0.83; *P*=.01), and the 1-month follow-up (between-group Cohen *d*=0.33, 95% CI −0.03 to 0.69; *P*=.07); thus, it is likely that high perceived academic functioning is a predictor of booster use. In contrast, between-group effects on the PHQ-9 did not emerge until after the booster was offered. Between-group effects were not significant for the GAD-7 (*P*=.21) or SDS (*P*=.61) at the 3-month follow-up. Table 3 includes additional details on the between-group effects.

**Table 3.** Between groups effect sizes (Cohen *d*) for primary and secondary outcomes based on MI[a] and booster factors using pooled imputations.

| Outcomes | After treatment, Cohen *d* (95% CI) | 1-month follow-up, Cohen *d* (95% CI) | 3-month follow-up, Cohen *d* (95% CI) |
|---|---|---|---|
| **MI** | | | |
| **Primary outcomes** | | | |
| PHQ-9[b] | 0.23 (−0.01 to 0.47) | 0.14 (−0.10 to 0.37) | 0.08 (−0.16 to 0.31) |
| GAD-7[c] | 0.25 (0.02 to 0.49) | 0.07 (−0.17 to 0.30) | 0.06 (−0.17 to 0.30) |
| PAF[d] | 0.05 (−0.18 to 0.29) | −0.04 (−0.27 to 0.20) | 0.09 (−0.15 to 0.32) |
| **Secondary outcomes** | | | |
| SDS[e] | 0.35 (0.12 to 0.59) | 0.08 (−0.16 to 0.31) | −0.02 (−0.25 to 0.22) |
| AUDIT[f] | 0.11 (−0.12 to 0.35) | —[g] | — |
| DUDIT[h] | 0.08 (−0.15 to 0.32) | — | — |
| **Booster** | | | |
| **Primary outcomes** | | | |
| PHQ-9 | −0.17 (−0.40 to 0.07) | −0.09 (−0.32 to 0.15) | 0.00 (−0.23 to 0.24) |
| GAD-7 | −0.24 (−0.48 to −0.01) | −0.16 (−0.39 to 0.08) | −0.09 (−0.33 to 0.14) |
| PAF | 0.10 (−0.13 to 0.34) | 0.08 (−0.15 to 0.32) | −0.01 (−0.25 to 0.23) |
| **Secondary outcomes** | | | |
| SDS | −0.05 (−0.28 to 0.19) | 0.00 (−0.24 to 0.23) | 0.03 (−0.20 to 0.27) |
| AUDIT | 0.00 (−0.23 to 0.24) | — | — |
| DUDIT | 0.20 (−0.03 to 0.44) | — | — |

[a] MI: motivational interviewing.

[b] PHQ-9: 9-item Patient Health Questionnaire.

[c] GAD-7: 7-item Generalized Anxiety Disorder.

[d] PAF: Perceptions of Academic Functioning.

[e] SDS: Sheehan Disability Scale.

[f] AUDIT: Alcohol Use Disorder Identification Test.

[g] The AUDIT and DUDIT were only administered before treatment and after treatment; thus, data are not available for the percentage change and effect sizes at the 1-month and 3-month follow-ups.

[h] DUDIT: Drug Use Disorder Identification Test.

## Secondary Outcome Measures

Table 2 also includes details on the estimated marginal means, percentage reductions, and effect sizes for each of the secondary measures separated by factor. Multimedia Appendix 1 shows an overview of this information, separated into the 4 treatment conditions. Significant time effects were found for the SDS, regardless of factor. Table 3 summarizes the between-group effect sizes.

For MI, a small between-group effect was found after treatment, such that clients who received MI had larger improvements on the SDS than clients who did not receive MI (between-group Cohen *d*=0.35, 95% CI 0.12-0.59). At the 1-month (Cohen *d*=−0.24 to 0.23) and 3-month follow-ups (Cohen *d*=−0.20 to 0.27), these differences were no longer present, and there were large within-group effect sizes for improvements on the SDS, regardless of factor (Cohen *d*=1.02-1.25) and the 3-month follow-up (Cohen *d*=0.97-1.18). No main effect was found for MI for the AUDIT (*P*=.35) or DUDIT (*P*=.49) after treatment, and these measures were not administered during follow-up.

No main effects for randomization to the booster were found for the SDS (*P*=.78). Similarly, the subanalysis comparing those who accessed the booster and those who did not access the booster failed to find group differences (*P*=.61). The AUDIT and DUDIT were not administered at the 3-month follow-up.

## Clinical Significance

After treatment, 47.7% (132/277) of all clients met the criteria for reliable recovery, 60.3% (167/277) met the criteria for reliable improvement, 1.9% (5/277) met the criteria for deterioration, and 37.9% (105/277) met the criteria for no change on the PHQ-9. For the GAD-7, the rate of reliable recovery was 56.6% (157/277), the rate of reliable improvement was 75.5% (209/277), the rate of deterioration was 2.2% (6/277), and the rate of no change was 22% (61/277). At all time points, no significant main effects were found for MI or booster (*P*=.13-.99).

### Treatment Engagement

Of the clients in one of the MI conditions, 88.9% (120/135) completed the MI exercises and started lesson 1 (MI: 60/67, 90%; MI+booster: 60/68, 88%). Overall, 66.8% (183/277) of the clients accessed at least three of the four lessons, and 54.2% (150/277) accessed all 4 lessons within the 5 weeks of treatment. Of the clients who were randomized to booster, 30.9% (43/139) accessed the booster. Across conditions, clients logged in an average of 12.95 (SD 9.15) times, received an average of 5.23 (SD 0.83) emails from their therapists, and sent an average of 1.98 (SD 1.71) emails to their therapists. There was an average of 29.05 (SD 19.16) days between the clients' enrollment date and their last log-in to the treatment portal. No main effects were found for MI on any measure of treatment engagement ($P$=.11-.75). Similarly, the main effects for the booster were not significant ($P$=.21-.95). Subanalyses comparing those who accessed the booster and those who did not found that clients who accessed the booster had a greater number of days between enrollment and their last visit to the treatment portal (mean 53.19, SD 15.70 vs mean 34.49, SD 20.65; $F_{1,267}$=3.41; $P$=.07) and received a greater number of phone calls from their therapist (mean 0.51, SD 0.87 vs mean 0.44, SD 0.51; $F_{1,273}$=3.21; $P$=.07). Therapists spent an average of 110.55 (SD 43.66) minutes monitoring client progress and supporting each client, with no significant differences found for mean time per client ($F_{3,274}$=11.14; $P$=.33).

### Treatment Satisfaction

Clients reported high rates of satisfaction overall, with 82.3% (158/192), 85.5% (165/193), and 84.5% (163/193) reporting that they were satisfied or very satisfied with the treatment, the treatment platform, and the lessons and do-it-yourself guides, respectively. Most clients felt that the treatment was worth their time (171/192, 89.1%) and that they would recommend it to a friend (176/193, 91.2%). Furthermore, 82.9% (160/193) of clients reported that their motivation to seek help if needed in the future either increased or greatly increased, and 76.2% (147/193) felt that their confidence in their ability to manage their symptoms either increased or greatly increased. No significant differences were found between treatment conditions on any of the treatment satisfaction measures ($P$=.37-.83).

## Discussion

Although ICBT is an effective treatment option for postsecondary students experiencing symptoms of anxiety or depression [11], there are concerns about treatment completion and outcomes in this population. This study extends previous work on ICBT for postsecondary students by examining the inclusion of pretreatment MI exercises and a self-guided booster lesson offered 1 month after treatment.

### Impact of MI Exercises

Some benefit was found for the inclusion of pretreatment MI on symptoms of depression, anxiety, and overall functioning after treatment. Clients who were randomized to one of the MI conditions reported larger improvements in symptoms of depression, anxiety, and overall functioning from before treatment to after treatment than those of clients who were not assigned MI. No benefit for MI was found at either the 1-month or 3-month follow-up; thus, it appears that pretreatment MI may only result in temporary benefits compared with ICBT without MI. MI did not contribute to higher rates of treatment completion or greater engagement (ie, more log-ins to the website, more days enrolled in the course, or more client messages sent to therapists). Findings from this trial replicate those of a previous trial that examined pretreatment MI before an 8-week ICBT program [19] with 1 exception—we found some evidence for pretreatment MI contributing to greater symptom improvement for depression, anxiety, and overall functioning after treatment, whereas the previous trial found no benefits.

An explanation for why the MI exercises improved some outcomes despite no observable increase in treatment engagement is that the MI exercises helped elicit more change talk from the clients. Change talk was not examined directly in this study; however, a previous trial found that clients who completed pretreatment MI exercises included more change talk statements in messages with therapists than those who did not complete the exercises, despite no differences in treatment completion rates between the groups [19]. Other studies have described how the inclusion of MI in ICBT can lower client resistance to treatment [49]. Clients who completed the MI exercises may have been more engaged with the lessons and homework activities they completed, although they did not complete more lessons overall.

Pretreatment MI may not have led to higher treatment completion rates in both this trial and that of Soucy et al [19] as clients already have relatively high mean pretreatment motivation (CQ-3 scores in the study by Soucy et al [19] 25.44-25.59; CQ-3 scores in this trial 23.39-24.35). It had been hypothesized that students experience low levels of motivation in ICBT; however, the findings of this trial suggest otherwise, and it may be that the MI exercises are not relevant for many clients. Within the literature on face-to-face CBT, it has been reported that integrating MI throughout CBT can lead to higher rates of recovery in GAD at the 1-year follow-up than with CBT alone [50]. Thus, it may be worthwhile for future research to explore the integration of MI throughout the course of ICBT and the targeting of MI among less-motivated clients.

### Impact of Booster

The inclusion of a self-guided booster lesson in ICBT for postsecondary students has not been previously examined; thus, no hypotheses were made regarding the proportion of clients who would make use of a booster. Overall, there were no significant differences between those assigned to the booster and those who were not. The lack of differences is likely, in part, related to the low use of the booster lesson. Booster use in this study was lower (43/134, 31.9%) than that in previous trials of boosters in ICBT (32/47, 68% in the study by Andersson et al [23] and 114/223, 51.5% in the study by Hadjistavropoulos et al [24]), although both these trials included therapist support during the booster, which may have been more favorable to clients than a self-guided booster lesson. There was evidence for lower symptoms of depression at the 3-month follow-up among clients who accessed the booster than among those who did not access the booster, which was in contrast with the findings of a recent trial of a therapist-assisted booster following ICBT [24]. It is possible that clients who felt that they were managing their academic studies well were more likely to believe that they had the time to review the booster lesson. In addition, clients who benefited from the first 5 weeks of treatment might have been more likely to access the booster, as it has been found that initial success with an intervention predicts booster outcomes [51]. Future research should explore the impact of boosters under varying conditions (eg, different periods, contents, and levels of support).

### Overall Outcomes

Although findings related to MI and the booster condition were limited and completion rates were slightly above 50%, across the treatment conditions, clients experienced large reductions in both depression (Cohen $d$=1.25-1.67) and anxiety (Cohen $d$=1.42-2.01) after treatment, replicating past findings on the *UniWellbeing* course in another context. The slightly larger effect sizes in this trial may be attributed to the fact that the clients in this trial had higher symptom severity before treatment. It is possible that clients in our trial experienced an exacerbation of pre-existing symptoms of depression and anxiety as a result of the COVID-19 pandemic [52], which may have contributed to their slightly higher mean scores on the PHQ-9 and GAD-7 before treatment. Clients who start ICBT with more severe symptoms have greater symptom improvement than those with less severe symptoms [13]. Furthermore, Dear et al [31] reported on subanalyses of clients who started with moderate to severe scores on the PHQ-9 and GAD-7 and found larger effect sizes for depression (Cohen $d$=1.42-1.97) and anxiety (Cohen $d$=1.93-2.13) among these clients.

Treatment completion rates were similar between this study (150/277, 54.1%) and a previous trial (59%) [31] and were even more similar when comparing this study with the clients who self-referred (53%) to the *UniWellbeing* course in the previous trial of *UniWellbeing* [31]. In this trial, we found that clients logged in more days on average than in the previous trial (12.95 vs 8.70 log-ins) [31]. The previous trial did not report on the average number of days that clients were enrolled in the course or the average number of messages that clients sent to therapists; therefore, a direct comparison cannot be made between the 2 trials.

Previous trials of the *UniWellbeing* course [30,31] did not include subjective measures of academic functioning; therefore, the inclusion of the PAF in this trial is a unique contribution. Although it was hypothesized that there would be moderate effects for improvements on the PAF, we found small effects from pretreatment to posttreatment across the 4 treatment conditions. The studies included in the meta-analysis conducted by Harrer et al [11] either used measures of overall functioning or relied on a single objective indicator of academic functioning, such as grade-point average. Therefore, a direct comparison between the findings on subjective academic functioning in this trial and the findings of Harrer et al [11] is not possible. The finding that clients only experienced small improvements in perceived academic functioning should be considered within the context of the COVID-19 pandemic. Students faced considerable uncertainty about classes, lectures, examination formats, and the future of their academic careers [53], and this may factor into their overall ability to cope with academic pressures while completing ICBT.

It is challenging to compare the findings of this trial with the overall effects from meta-analyses of ICBT for postsecondary students [11] because of the heterogeneity of the studies included. Studies varied substantially in terms of guidance (eg, self-guided vs therapist-assisted), recruitment (eg, psychology participant pool vs clinical sample), and treatment modality (eg, website vs app). Overall, the effect sizes for improvements in depression and anxiety appeared larger in this trial than in the meta-analysis [11] and may be explained by the inclusion of weekly therapist support in this trial [54] and the requirement that clients have at least mild symptoms at intake [13].

Although the *UniWellbeing* course has been compared with a wait-list control in the past [30], a control group was not included in this study, and it is likely that some of the improvements in symptoms were because of a regression to the mean. It has been reported that 20% of students with major depressive disorder experience remission at the end of a 9-week observation period in the absence of treatment [55]; therefore, it would be expected that a proportion of students in this trial may have improved without ICBT.

### Limitations and Future Directions

There were several limitations to this study, which can inform future trials of ICBT for postsecondary students. One of the limitations was that the MI component was only offered before treatment, which may not have been the most beneficial time to offer MI to clients as clients may be starting treatment with high levels of motivation. In future trials of the *UniWellbeing* course, clients could be offered an MI resource that they could access at any point during the course, as opposed to before treatment. If therapists note client disengagement, ambivalence, or resistance, they could direct the client to this resource, similar to how therapists direct clients to additional resources (eg, sleep and assertive communication) as part of their practice when delivering ICBT. Soucy et al [19] found that the MI exercises resulted in an increase in client motivation and *change talk*; thus, client use of an MI resource during ICBT could facilitate greater treatment completion when offered at a more appropriate time.

Future studies could also examine a *blended* version of the *UniWellbeing* course, whereby the therapist contact remains unchanged for the 4 core lessons, but the client has the opportunity to schedule an appointment with their therapist via telephone or secure video call to review the MI exercises. A telephone or video call would enable the client to receive direct feedback from their therapist and would provide opportunities for the therapist to respond in the “spirit” of MI (ie, emphasizing collaboration with the client, evoking change, and emphasizing the autonomy of the client to initiate change) [33]. A longer follow-up period (eg, 6 or 12 months after treatment) would allow for a better understanding of the long-term effects of MI on symptom reduction. It is also possible that treatment noncompletion was not related to motivation but instead to students having preferences for different treatment doses. Future studies could offer students the choice between a brief, standard

(5-week), or extended version of the ICBT course to align with student preferences.

A limitation of the booster lesson was that it was offered only 1 month after treatment completion. Some clients may not have felt that they needed a booster lesson soon after treatment, which likely contributed to the low overall uptake of the booster. Among those who used the booster, there was preliminary evidence suggesting that the booster was associated with larger reductions in depression at the 3-month follow-up; however, these subanalyses were underpowered because of low uptake. Furthermore, we are unable to comment on the longer-term impacts of the booster lesson on symptoms of depression and anxiety, as well as subjective academic functioning, given that the final outcome measures were administered at the 3-month follow-up. Andersson et al [23] found that boosters in ICBT can help reduce relapse rates of obsessive-compulsive disorder up to at least the 1-year follow-up; therefore, it would be worthwhile to examine the impact of a self-guided booster lesson over a longer follow-up period than used in this trial.

It should be noted that all the clients enrolled in this trial started treatment during the COVID-19 pandemic. It is possible that clients experienced regression to the mean in terms of their symptom severity as they became accustomed to COVID-19 public health restrictions and the impact on their academic studies. As this is the first trial of the *UniWellbeing* course in Saskatchewan, we do not have a comparison sample to use as a benchmark. Only 16.2% (45/277) of our sample identified as men; thus, future studies should attempt to recruit a more balanced representation of genders. A further limitation is that a modified ITT approach was used, which eliminated any data from the 10.1% of the clients who did not start treatment. Finally, the PAF was designed for this study and requires further validation beyond internal consistency in future research.

## Strengths

A notable strength of this study is that we were able to replicate the findings of a previous trial on the *UniWellbeing* course [31] in terms of large reductions in symptoms of depression and anxiety that were maintained at the 3-month follow-up. This replication is important as it provides evidence for the generalizability of the initial findings in a different country (ie, Canada) and within the context of the COVID-19 pandemic. An additional strength is the use of a factorial design, which allowed for the concurrent examination of both pretreatment MI and a self-guided booster lesson within a single trial. To the best of our knowledge, this combination of factors has not been examined in previous trials of ICBT for postsecondary students. This trial also provides useful information about students' interest in, and uptake of, a self-guided booster lesson following ICBT.

## Conclusions

The findings from this factorial trial provide evidence for the efficacy of a 5-week transdiagnostic ICBT course for postsecondary students. Large effect sizes were found for reductions in symptoms of depression and anxiety, and small effect sizes were found for improvements in perceived academic functioning, with changes maintained up to the 3-month follow-up. There was some evidence for the benefit of pretreatment MI in improving depression, anxiety, and disability outcomes after treatment; however, no benefit was found for treatment completion or engagement. No main effects were found for the inclusion of a booster. However, although the booster was used by less than one-third of clients, there was some evidence for improved depression outcomes at the 3-month follow-up among booster users. Further research could explore whether it is possible to optimize ICBT for postsecondary populations by using variations of MI and booster lessons.

## Acknowledgments

The authors wish to acknowledge the clients, patient partners, screeners, therapists, research staff, and web developers from the Online Therapy Unit at the University of Regina, as well as the therapists from the Saskatchewan Health Authority for their contributions to this project.

This research was conducted by the Online Therapy Unit, which receives funding from the Saskatchewan Ministry of Health to provide internet-delivered cognitive behavioral therapy to residents of Saskatchewan. The funders had no involvement in the design of the paper and the collection, analysis, or interpretation of the data. NT and BD were funded by the Australian government to operate the MindSpot Clinic. NT and BD developed the *UniWellbeing* course but derive no financial benefit from it.

## Data Availability

Data are not available because of confidentiality concerns.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Estimated marginal means, 95% CIs, percentage changes, and effect sizes (Cohen *d*) for primary and secondary outcomes by treatment condition using pooled imputations after treatment and at the 1-month and 3-month follow-ups.

[DOCX File , 24 KB - jmir_v24i9e40637_app1.docx ]

## References

1. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]
2. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
3. 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]
4. Wallace KF, Putnam NI, Chow E, Fernandes M, Clary KM, Goff SL. College students' experiences early in the COVID-19 pandemic: applications for ongoing support. J Am Coll Health 2021 Aug 17:1-10. [doi: 10.1080/07448481.2021.1954011] [Medline: 34403622]
5. Lederer AM, Hoban MT, Lipson SK, Zhou S, Eisenberg D. More than inconvenienced: the unique needs of U.S. college students during the COVID-19 pandemic. Health Educ Behav 2021 Feb;48(1):14-19 [FREE Full text] [doi: 10.1177/1090198120969372] [Medline: 33131325]
6. Vaterlaus JM, Shaffer T, Pulsipher L. College student interpersonal and institutional relationships during the COVID-19 pandemic: a qualitative exploratory study. Soc Sci J 2021 Jul 16:1-14. [doi: 10.1080/03623319.2021.1949553]
7. Lindberg LD, Bell DL, Kantor LM. The sexual and reproductive health of adolescents and young adults during the COVID-19 pandemic. Perspect Sex Reprod Health 2020 Jul;52(2):75-79 [FREE Full text] [doi: 10.1363/psrh.12151] [Medline: 32537858]
8. Ebert DD, Mortier P, Kaehlke F, Bruffaerts R, Baumeister H, Auerbach RP, WHO World Mental Health-International College Student Initiative collaborators. Barriers of mental health treatment utilization among first-year college students: first cross-national results from the WHO World Mental Health International College Student Initiative. Int J Methods Psychiatr Res 2019 Jun;28(2):e1782 [FREE Full text] [doi: 10.1002/mpr.1782] [Medline: 31069905]
9. Bruffaerts R, Mortier P, Auerbach RP, Alonso J, Hermosillo De la Torre AE, Cuijpers P, WHO WMH-ICS Collaborators. Lifetime and 12-month treatment for mental disorders and suicidal thoughts and behaviors among first year college students. Int J Methods Psychiatr Res 2019 Jun;28(2):e1764 [FREE Full text] [doi: 10.1002/mpr.1764] [Medline: 30663193]
10. Etzelmueller A, Vis C, Karyotaki E, Baumeister H, Titov N, Berking M, et al. Effects of Internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. J Med Internet Res 2020 Aug 31;22(8):e18100 [FREE Full text] [doi: 10.2196/18100] [Medline: 32865497]
11. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
12. Davies EB, Morriss R, Glazebrook C. Computer-delivered and Web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
13. Edmonds M, Hadjistavropoulos HD, Schneider LH, Dear BF, Titov N. Who benefits most from therapist-assisted internet-delivered cognitive behaviour therapy in clinical practice? Predictors of symptom change and dropout. J Anxiety Disord 2018 Mar;54:24-32 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.003] [Medline: 29421369]
14. Pedrelli P, Nyer M, Yeung A, Zulauf C, Wilens T. College students: mental health problems and treatment considerations. Acad Psychiatry 2015 Oct;39(5):503-511 [FREE Full text] [doi: 10.1007/s40596-014-0205-9] [Medline: 25142250]
15. Randall CL, McNeil DW. Motivational interviewing as an adjunct to cognitive behavior therapy for anxiety disorders: a critical review of the literature. Cogn Behav Pract 2017 Aug;24(3):296-311 [FREE Full text] [doi: 10.1016/j.cbpra.2016.05.003] [Medline: 28871216]
16. Marker I, Norton PJ. The efficacy of incorporating motivational interviewing to cognitive behavior therapy for anxiety disorders: a review and meta-analysis. Clin Psychol Rev 2018 Jun;62:1-10. [doi: 10.1016/j.cpr.2018.04.004] [Medline: 29727863]
17. Samson JE, Tanner-Smith EE. Single-session alcohol interventions for heavy drinking college students: a systematic review and meta-analysis. J Stud Alcohol Drugs 2015 Jul;76(4):530-543 [FREE Full text] [doi: 10.15288/jsad.2015.76.530] [Medline: 26098028]
18. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312 [FREE Full text] [Medline: 15826439]
19. Soucy JN, Hadjistavropoulos HD, Karin E, Dear BF, Titov N. Brief online motivational interviewing pre-treatment intervention for enhancing internet-delivered cognitive behaviour therapy: a randomized controlled trial. Internet Interv 2021 Sep;25:100394 [FREE Full text] [doi: 10.1016/j.invent.2021.100394] [Medline: 33996510]
20. Titov N, Andrews G, Schwencke G, Robinson E, Peters L, Spence J. Randomized controlled trial of Internet cognitive behavioural treatment for social phobia with and without motivational enhancement strategies. Aust N Z J Psychiatry 2010 Oct;44(10):938-945. [doi: 10.3109/00048674.2010.493859] [Medline: 20932208]

21. Whisman MA. The efficacy of booster maintenance sessions in behavior therapy: review and methodological critique. Clin Psychol Rev 1990 Jan;10(2):155-170. [doi: 10.1016/0272-7358(90)90055-F]
22. Gearing RE, Schwalbe CS, Lee R, Hoagwood KE. The effectiveness of booster sessions in CBT treatment for child and adolescent mood and anxiety disorders. Depress Anxiety 2013 Sep;30(9):800-808. [doi: 10.1002/da.22118] [Medline: 23596102]
23. Andersson E, Steneby S, Karlsson K, Ljótsson B, Hedman E, Enander J, et al. Long-term efficacy of Internet-based cognitive behavior therapy for obsessive-compulsive disorder with or without booster: a randomized controlled trial. Psychol Med 2014 Oct;44(13):2877-2887. [doi: 10.1017/S0033291714000543] [Medline: 25066102]
24. Hadjistavropoulos HD, Peynenburg V, Thiessen DL, Nugent M, Karin E, Dear BF, et al. A randomized factorial trial of internet-delivered cognitive behavioural therapy: an 8-week program with or without extended support and booster lesson. Internet Interv 2022 Mar;27:100499 [FREE Full text] [doi: 10.1016/j.invent.2022.100499] [Medline: 35198410]
25. Gibbons CR, Stirman SW, Derubeis RJ, Newman CF, Beck AT. Research setting versus clinic setting: which produces better outcomes in cognitive therapy for depression? Cognit Ther Res 2013 Jun 01;37(3):605-612 [FREE Full text] [doi: 10.1007/s10608-012-9499-7] [Medline: 23710102]
26. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]
27. Watkins E, Newbold A, Tester-Jones M, Javaid M, Cadman J, Collins LM, et al. Implementing multifactorial psychotherapy research in online virtual environments (IMPROVE-2): study protocol for a phase III trial of the MOST randomized component selection method for internet cognitive-behavioural therapy for depression. BMC Psychiatry 2016 Oct 06;16(1):345 [FREE Full text] [doi: 10.1186/s12888-016-1054-8] [Medline: 27716200]
28. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
29. Spitzer RL, Kroenke K, Williams JB, 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]
30. Mullin A, Dear BF, Karin E, Wootton BM, Staples LG, Johnston L, et al. The UniWellbeing course: a randomised controlled trial of a transdiagnostic internet-delivered cognitive behavioural therapy (CBT) programme for university students with symptoms of anxiety and depression. Internet Interv 2015 May;2(2):128-136 [FREE Full text] [doi: 10.1016/j.invent.2015.02.002]
31. Dear BF, Johnson B, Singh A, Wilkes B, Brkic T, Gupta R, et al. Examining an Internet-delivered intervention for anxiety and depression when delivered as a part of routine care for university students: a phase IV trial. J Affect Disord 2019 Sep 01;256:567-577 [FREE Full text] [doi: 10.1016/j.jad.2019.06.044] [Medline: 31280082]
32. Beck CD, Soucy JN, Hadjistavropoulos HD. Mixed-method evaluation of an online motivational intervention as a pre-treatment to internet-delivered cognitive behaviour therapy: immediate benefits and user feedback. Internet Interv 2020 Apr;20:100311 [FREE Full text] [doi: 10.1016/j.invent.2020.100311] [Medline: 32154118]
33. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change. 3rd edition. New York, NY, USA: The Guilford Press; 2013.
34. Friederichs SA, Oenema A, Bolman C, Guyaux J, van Keulen HM, Lechner L. I Move: systematic development of a Web-based computer tailored physical activity intervention, based on motivational interviewing and self-determination theory. BMC Public Health 2014 Feb 28;14:212 [FREE Full text] [doi: 10.1186/1471-2458-14-212] [Medline: 24580802]
35. Hadjistavropoulos HD, Schneider LH, Klassen K, Dear BF, Titov N. Development and evaluation of a scale assessing therapist fidelity to guidelines for delivering therapist-assisted Internet-delivered cognitive behaviour therapy. Cogn Behav Ther 2018 Nov;47(6):447-461. [doi: 10.1080/16506073.2018.1457079] [Medline: 29693533]
36. Hadjistavropoulos HD, Williams J, Adlam K, Spice K, Nugent M, Owens KM, et al. Audit and feedback of therapist-assisted internet-delivered cognitive behaviour therapy within routine care: a quality improvement case study. Internet Interv 2020 Apr;20:100309 [FREE Full text] [doi: 10.1016/j.invent.2020.100309] [Medline: 32071887]
37. Sheehan DV. The Sheehan Disability Scales. The Anxiety Disease and How to Overcome It. New York, NY, USA: Charles Scribner and Sons; 1983.
38. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
39. World Health Organization, Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. The Alcohol Use Disorders Identification Test: Guidelines for use in primary care. 2nd edition. World Health Organization. 2001. URL: https://apps.who.int/iris/handle/10665/67205 [accessed 2022-01-30]
40. Berman AH, Bergman H, Palmstierna T, Schlyter F. Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample. Eur Addict Res 2005;11(1):22-31. [doi: 10.1159/000081413] [Medline: 15608468]
41. Miller WR, Johnson WR. A natural language screening measure for motivation to change. Addict Behav 2008 Sep;33(9):1177-1182. [doi: 10.1016/j.addbeh.2008.04.018] [Medline: 18558466]

42. Dear BF, Titov N, Schwencke G, Andrews G, Johnston L, Craske MG, et al. An open trial of a brief transdiagnostic internet treatment for anxiety and depression. Behav Res Ther 2011 Dec;49(12):830-837. [doi: 10.1016/j.brat.2011.09.007] [Medline: 21993153]
43. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
44. Karin E, Crane MF, Dear BF, Nielssen O, Heller GZ, Kayrouz R, et al. Predictors, outcomes, and statistical solutions of missing cases in Web-based psychotherapy: methodological replication and elaboration study. JMIR Ment Health 2021 Feb 05;8(2):e22700 [FREE Full text] [doi: 10.2196/22700] [Medline: 33544080]
45. Karin E, Dear BF, Heller GZ, Crane MF, Titov N. "Wish you were here": examining characteristics, outcomes, and statistical solutions for missing cases in Web-based psychotherapeutic trials. JMIR Ment Health 2018 Apr 19;5(2):e22 [FREE Full text] [doi: 10.2196/mental.8363] [Medline: 29674311]
46. Karin E, Dear BF, Heller GZ, Gandy M, Titov N. Measurement of symptom change following Web-based psychotherapy: statistical characteristics and analytical methods for measuring and interpreting change. JMIR Ment Health 2018 Jul 12;5(3):e10200 [FREE Full text] [doi: 10.2196/10200] [Medline: 30001999]
47. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided Internet-delivered treatment for generalized anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
48. Hadjistavropoulos HD, Nugent MM, Alberts NM, Staples L, Dear BF, Titov N. Transdiagnostic Internet-delivered cognitive behaviour therapy in Canada: an open trial comparing results of a specialized online clinic and nonspecialized community clinics. J Anxiety Disord 2016 Aug;42:19-29 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.006] [Medline: 27244278]
49. Aviram A, Westra HA. The impact of motivational interviewing on resistance in cognitive behavioural therapy for generalized anxiety disorder. Psychother Res 2011 Nov;21(6):698-708. [doi: 10.1080/10503307.2011.610832] [Medline: 21878034]
50. Westra HA, Constantino MJ, Antony MM. Integrating motivational interviewing with cognitive-behavioral therapy for severe generalized anxiety disorder: an allegiance-controlled randomized clinical trial. J Consult Clin Psychol 2016 Sep;84(9):768-782. [doi: 10.1037/ccp0000098] [Medline: 26985729]
51. Baker AL, Wilson PH. Cognitive-behavior therapy for depression: the effects of booster sessions on relapse. Behav Ther 1985 Sep;16(4):335-344. [doi: 10.1016/s0005-7894(85)80001-0]
52. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
53. Aristovnik A, Keržič D, Ravšelj D, Tomaževič N, Umek L. Impacts of the COVID-19 pandemic on life of higher education students: a global perspective. Sustainability 2020 Oct 13;12(20):8438. [doi: 10.3390/su12208438]
54. 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]
55. Kelly MA, Roberts JE, Bottonari KA. Non-treatment-related sudden gains in depression: the role of self-evaluation. Behav Res Ther 2007 Apr;45(4):737-747. [doi: 10.1016/j.brat.2006.06.008] [Medline: 16905116]

## Abbreviations

**AUDIT:** Alcohol Use Disorder Identification Test
**CBT:** cognitive behavioral therapy
**CQ-3:** 3-item Change Questionnaire
**DUDIT:** Drug Use Disorders Identification Test
**GAD:** generalized anxiety disorder
**GAD-7:** 7-item Generalized Anxiety Disorder
**ICBT:** internet-delivered cognitive behavioral therapy
**ITT:** intention-to-treat
**MI:** motivational interviewing
**PAF:** Perceptions of Academic Functioning
**PHQ-9:** 9-item Patient Health Questionnaire
**REDCap:** Research Electronic Data Capture
**SDS:** Sheehan Disability Scale

*Edited by G Eysenbach; submitted 29.06.22; peer-reviewed by D Eisenberg, E Ennis, J Sharp; comments to author 22.07.22; revised version received 28.07.22; accepted 12.08.22; published 07.09.22.*

*Please cite as:*
*Peynenburg V, Hadjistavropoulos H, Thiessen D, Titov N, Dear B*
*Internet-Delivered Cognitive Behavioral Therapy for Postsecondary Students: Randomized Factorial Trial for Examining Motivational Interviewing and Booster Lessons*
*J Med Internet Res 2022;24(9):e40637*
*URL: https://www.jmir.org/2022/9/e40637*
*doi:10.2196/40637*
*PMID:36069785*

Original Paper

# Web-Based Mindfulness-Based Interventions for Well-being: Randomized Comparative Effectiveness Trial

Louisa G Sylvia[1,2], PhD; Mitchell R Lunn[3], MAS, MD; Juno Obedin-Maliver[3], MAS, MPH, MD; Robert N McBurney[4], PhD; W Benjamin Nowell[5], MSW, PhD; Rachel L Nosheny[6], PhD; Richard A Mularski[7], MSHS, MCR, MD; Millie D Long[8], MPH, MD; Peter A Merkel[9], MPH, MD; Mark J Pletcher[6], MPH, MD; Roberta E Tovey[1], PhD; Christopher Scalchunes[10], MPA; Rebecca Sutphen[11], MD; Ann S Martin[12], MSc; Elizabeth J Horn[13], PhD; Megan O'Boyle[13], BA; Lisa Pitch[14], MA; Michael Seid[15], PhD; Susan Redline[16], MPH, MD; Sophie Greenebaum[1], BA; Nevita George[1], BSc; Noah J French[1], BA; Caylin M Faria[1], BSc; Nicha Puvanich[1], MSc; Dustin J Rabideau[17], PhD; Caitlin A Selvaggi[17], MSc; Chu Yu[17], MSc; Stephen V Faraone[18], PhD; Shilpa Venkatachalam[5], MPH, PhD; Debbe McCall[19], MBA; Sharon F Terry[20], MA; Thilo Deckersbach[1], PhD; Andrew A Nierenberg[1,2], MD

[1]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
[2]Harvard University, Cambridge, MA, United States
[3]Stanford University School of Medicine, Stanford, CA, United States
[4]Accelerated Cure Project for Multiple Sclerosis, Waltham, MA, United States
[5]Global Healthy Living Foundation, Upper Nyack, NY, United States
[6]University of California San Francisco School of Medicine, San Francisco, CA, United States
[7]Kaiser Permanente Center for Health Research Northwest Region, Portland, OR, United States
[8]University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, United States
[9]University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States
[10]Immune Deficiency Foundation, Towson, MD, United States
[11]University of South Florida, Tampa, FL, United States
[12]Parent Project Muscular Dystrophy, Washington, DC, United States
[13]Phelan-McDermid Syndrome Foundation, Osprey, FL, United States
[14]ImproveCareNow Inc., Burlington, VT, United States
[15]Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[16]Brigham and Women's Hospital, Boston, MA, United States
[17]Biostatistics, Massachusetts General Hospital, Boston, MA, United States
[18]State University of New York Upstate Medical University, Syracuse, NY, United States
[19]Health eHeart Alliance, San Francisco, CA, United States
[20]Genetic Alliance, Damascus, MD, United States

**Corresponding Author:**
Louisa G Sylvia, PhD
Department of Psychiatry
Massachusetts General Hospital
50 Staniford Street, Suite 580
Boston, MA, 02114
United States
Phone: 1 617 643 4804
Email: lsylvia2@mgh.harvard.edu

## Abstract

**Background:** Mindfulness can improve overall well-being by training individuals to focus on the present moment without judging their thoughts. However, it is unknown how much mindfulness practice and training are necessary to improve well-being.

**Objective:** The primary aim of this study was to determine whether a standard 8-session web-based mindfulness-based cognitive therapy (MBCT) program, compared with a brief 3-session mindfulness intervention, improved overall participant well-being.

In addition, we sought to explore whether the treatment effects differed based on the baseline characteristics of the participants (ie, moderators).

**Methods:** Participants were recruited from 17 patient-powered research networks, web-based communities of stakeholders interested in a common research area. Participants were randomized to either a standard 8-session MBCT or a brief 3-session mindfulness training intervention accessed on the web. The participants were followed for 12 weeks. The primary outcome of the study was well-being, as measured by the World Health Organization—Five Well-Being Index. We hypothesized that MBCT would be superior to a brief mindfulness training.

**Results:** We randomized 4411 participants, 3873 (87.80%) of whom were White and 3547 (80.41%) of female sex assigned at birth. The mean baseline World Health Organization—Five Well-Being Index score was 50.3 (SD 20.7). The average self-reported well-being in each group increased over the intervention period (baseline to 8 weeks; model-based slope for the MBCT group: 0.78, 95% CI 0.63-0.93, and brief mindfulness group: 0.76, 95% CI 0.60-0.91) as well as the full study period (ie, intervention plus follow-up; baseline to 20 weeks; model-based slope for MBCT group: 0.41, 95% CI 0.34-0.48; and brief mindfulness group: 0.33, 95% CI 0.26-0.40). Changes in self-reported well-being were not significantly different between MBCT and brief mindfulness during the intervention period (model-based difference in slopes: −0.02, 95% CI −0.24 to 0.19; *P*=.80) or during the intervention period plus 12-week follow-up (−0.08, 95% CI −0.18 to 0.02; *P*=.10). During the intervention period, younger participants (*P*=.05) and participants who completed a higher percentage of intervention sessions (*P*=.005) experienced greater improvements in well-being across both interventions, with effects that were stronger for participants in the MBCT condition. Attrition was high (ie, 2142/4411, 48.56%), which is an important limitation of this study.

**Conclusions:** Standard MBCT improved well-being but was not superior to a brief mindfulness intervention. This finding suggests that shorter mindfulness programs could yield important benefits across the general population of individuals with various medical conditions. Younger people and participants who completed more intervention sessions reported greater improvements in well-being, an effect that was more pronounced for participants in the MBCT condition. This finding suggests that standard MBCT may be a better choice for younger people as well as treatment-adherent individuals.

**Trial Registration:** ClinicalTrials.gov NCT03844321; https://clinicaltrials.gov/ct2/show/NCT03844321

***(J Med Internet Res 2022;24(9):e35620)*** doi:10.2196/35620

**KEYWORDS**

mindfulness; well-being; web; control trial; clinical trial; cognitive therapy; intervention; mental health; mindful; eHealth; mobile phone

## *Introduction*

### Background

Many people with chronic diseases, and their caregivers, experience stress and decreased well-being. Data from the World Happiness Report [1] and Gallup Index for community well-being [2] suggest that overall life satisfaction and well-being have declined since 2017. In 2011, the American Psychological Association conducted the “Stress in America” survey of more than 1000 individuals with chronic illnesses (eg, depression, type 2 diabetes, obesity, or heart disease [3]). Owing to stress, 44% reported having trouble sleeping, 39% reported overeating, and 29% reported skipping meals. Stress and reduced well-being are common among caregivers. One in every 2 caregivers reports being *overwhelmed* with their caregiving responsibilities and feeling more stressed than the general population [3]. Caregivers are at risk of increased premature mortality, coronary heart disease, and stroke, particularly under conditions of high stress. Thus, it is imperative for people with chronic illnesses or their caregivers to develop methods to reduce stress to improve their overall well-being.

One of the most acceptable and effective interventions for improving one’s overall well-being are mindfulness-based treatments. Mindfulness improves well-being by cultivating awareness and being present in the moment [4]. The efficacy of mindfulness-based treatments has been well documented for a wide variety of psychiatric and medical problems, such as depression, anxiety, stress, alcohol and drug abuse, and pain [5]. Most empirically tested forms of mindfulness, such as mindfulness-based stress reduction (MBSR) or mindfulness-based cognitive therapy (MBCT), include a curriculum of guided meditation exercises taught in sequence over 2 to 3 months in weekly sessions (8-12 sessions as well as regular homework practice). However, the minimally effective *dose* of mindfulness is unknown.

One reasonable hypothesis is that more sessions and more time spent on practice would lead to larger treatment effects. For example, a dose-effect study found that more sessions of mindfulness were associated with greater improvement (ie, 30% improvement after 2 sessions, 41% after 4 sessions, 53% after 8 sessions, and 75% after 26 sessions) [6]. More generally, a study examining the session-by-session outcomes of nonspecific psychotherapy (N=6375 treated in 26 different centers, 94% from college counseling centers [4]) found that in session 3, a total of 32% patients had clinically significant improvement, and by session 8, a total of 50% participants had clinical improvement. However, both studies found a negatively accelerated dose-effect curve, such that the rate of improvement decreased with more sessions. Thus, a brief intervention could be sufficient to improve outcomes.

In particular, mindfulness-based interventions with only a few sessions (range 1-6) have been shown to increase mindfulness and decrease anxiety, depression, negative thoughts, headaches, and pain [7-12]. One study randomized healthy adults (N=120) to 1 session of mindfulness or an attention-only group and found greater reductions in stress as well as buffered physiological responses during social stress in the mindfulness group [13]. In patient cohort studies, patients with cardiac disease (N=114) were randomized to a 4-session mindfulness group or to a self-help control group, with results demonstrating that the intervention group yielded better outcomes on quality of life, anxiety, depression, and perceived stress, which were partially or fully mediated by an increase in mindfulness [14]. Therefore, very brief (1-3 sessions) mindfulness interventions may be effective in reducing stress and increasing well-being.

### Objectives

The primary aim of this study was to determine whether a standard 8-session web-based MBCT program compared with a brief 3-session mindfulness intervention improved well-being. In addition, we sought to explore whether the treatment effects differed based on the baseline characteristics of the participants (ie, moderators). Given that standard MBCT is the longer and more comprehensive intervention, we hypothesized that standard MBCT would be superior to a brief, 3-session mindfulness program in increasing well-being, quality of life, and functioning as well as decreasing stress, anxiety, and depression. The primary outcome measure was the World Health Organization—Five Well-Being Index (WHO-5) well-being index from baseline to 8 weeks as well as baseline to 20 weeks. There were no specific directional hypotheses for the moderator analyses.

## Methods

### Study Overview

We compared the effectiveness of an 8-session MBCT program and a 3-session brief mindfulness intervention to improve overall well-being. Adults (aged ≥18 years) from 17 web-based patient-powered research networks (PPRNs)—web-based people-centered organizations that focus on specific conditions and community interests through comparative effectiveness studies [15]—who were able to read and understand English and participate in mindfulness exercises were eligible to participate. Participants were either members of these 17 PPRNs or their families. The 17 PPRNs were assembled to conduct a demonstration project as part of the Patient-Centered Clinical Research Network. Patient-Centered Clinical Research Network consists of PPRNs, as well as clinical data research networks, with the intent of improving research by creating a national resource of health data, research expertise, and stakeholder experience. The PPRNs recruited for this study were based in the United States and represent a wide range of conditions and special populations (eg, people with arthritis, mood disorders, and Alzheimer disease and lesbian, gay, bisexual, transgender, queer, or questioning people); thus, the population of interest for this study was extraordinarily broad, with all adults belonging to the special populations represented and their caregivers. The study was conducted in collaboration with stakeholders (ie, patients, clinicians, advocates, researchers, caregivers, content experts in mindfulness, and patient-centered research) from several of the study's PPRNs. Stakeholders collaborated to determine the study's primary outcome, overall well-being, and identify the secondary outcomes of importance. The stakeholders oversaw study development, including testing and providing feedback on the web-based platform, and advised on enrollment strategies and the dissemination of study results.

### Study Platform

The study was hosted on a web-based platform developed by the same team that created IMoodNetwork [16], a PPRN at Massachusetts General Hospital for individuals with mood disorders. Following electronic consent, each participant completed a set of questionnaires that consisted of demographic, medical, and psychiatric history; history of mindfulness practice; and their role (ie, a member of the PPRN, or family member or caregiver of the PPRN member).

Eligible participants were randomly assigned to the 8-session web-based MBCT program or a brief web-based 3-session mindfulness program. The programs consisted of individual web-based modules comprising videos and activities delivered in short, digestible sections. The layout of the intervention material was optimized, or adapted, to the size of users' screens, and participants could therefore complete intervention and assessment sessions on a computer, tablet, or smartphone.

Randomization was performed using a stratified block randomization technique with a block size of 4 to maintain an even distribution across each PPRN. Randomization was executed using the MoodNetwork platform, which was programmed by the Massachusetts General Hospital study staff. Participants were not blinded to their randomization and were informed during the informed consent process that they would be assigned to a standard 8-week mindfulness program or a brief 3-week mindfulness program. All participants followed the same assessment schedule despite having different intervention schedules; thus, the active phase for assessments (weeks 0-8) and follow-up period for assessments (weeks 9-20) were the same for the brief mindfulness group and MBCT. During the active phase, assessments were performed every 2 weeks (weeks 0, 2, 4, 6, and 8), and during the follow-up period, assessments were performed every 2 months (weeks 12, 16, and 20).

The participants were prompted to return to the study platform to complete their activities and assessments via weekly email reminders. At the end of each intervention session, participants were instructed to practice mindfulness activities on most days; however, we did not gather data on how long the participants spent practicing the activities. The participants were entered into a raffle, and 5 participants were randomly selected to receive a US $200 Visa gift card.

### Interventions

#### *Standard MBCT Intervention*

The 8-session, standard MBCT program was based on the manual developed by Segal et al [17], which has been adapted for a wide variety of psychiatric disorders and medical conditions [18] as a web-based version with good efficacy [8]. Participants completed a structured curriculum of guided meditation exercises with 1 session per week for 8 weeks (eg, mindfulness of the breath, mindfulness of breath and body, mindfulness of thoughts and feelings, and open or choice-less awareness). Over the course of these exercises, participants learned to adopt an observing, accepting stance (mindfulness) toward difficult thoughts, feelings, and bodily sensations. Participants also learned to bring mindfulness to everyday situations and practice how to recognize and disengage from negative, ruminative thoughts.

### Brief Mindfulness Intervention

The 3-session brief mindfulness program was based on the work of Zeidan et al [9,10,19] and was also adapted to a web-based platform for this study. This brief mindfulness intervention has been shown to be more effective than sham meditation in reducing negative mood, depression, and fatigue [11]. Participants completed 1 session per week for 3 weeks, focusing on teaching a single breath-awareness meditation exercise during which participants learned to focus on the flow of their breath as well as *letting thoughts go* by bringing their attention back to the sensations of the breath. Participants received guidance on how to implement this skill during their daily lives and in stressful situations.

Both intervention groups were assigned mindfulness exercises to practice between sessions. They were also taught the core aspects of mindfulness (ie, adopting an observing, accepting a stance toward difficult thoughts, feelings, and bodily sensations). Participants in both groups were not able to perform more than one intervention session per week; however, they could access material from the previous weeks at any time.

### Study Assessments

Participants completed self-reported assessments at 8 time points (Multimedia Appendix 1, Table S3 provides a full schedule of study assessments). Assessments were available for 1 week and then were automatically closed after the due date passed.

### WHO-5 Well-Being Index (Primary Outcome)

This 5-item measure assesses well-being over the course of the prior 2 weeks (eg, “I have felt cheerful and in good spirits” or “I woke up feeling fresh and rested”) [20]. Participants rated how often they experienced each item on a scale from 0 (*at no time*) to 5 (*all of the time*). A score is computed by multiplying the total score by 4 (ranging from 0 to 100), with higher scores reflecting increased well-being. Participants completed this measure during each assessment period. A change of at least 10 points is estimated to be clinically meaningful [20,21].

### Demographics

Demographic variables including age, gender identity, sex assigned at birth, race, ethnicity, sexual orientation, marital status, employment status, and education history were measured at baseline.

### Medical and Psychiatric History

At baseline, participants were asked 2 questions: (1) “Do you have a history of any medical problem?” (response: yes or no); (2) “Do you have a history of any psychiatric illness?” (response: yes or no). If participants selected “yes” for having a history of a medical or psychiatric problem, they were asked to state the conditions for which they received treatment.

### Perceived Stress Scale

This 10-item measure evaluates an individual’s experience of stress in the past month (eg, “In the last month, how often have you been upset because of something that happened unexpectedly?” or “how often have you felt nervous and ‘stressed’?”) [5]. Participants rated how often they experienced these feelings and thoughts from 0 (*never*) to 4 (*very often*). Participants completed this measure during each assessment period. To our knowledge, estimates of minimum clinically important differences are not available.

### Patient-Reported Outcomes Measurement Information System: Emotional Distress-Depression Short Form

This questionnaire is an 8-item assessment of perceived depressive symptoms over the past week [22]. Participants rated how often they had experienced each item on a scale from 1 (*never*) to 5 (*always*). Participants completed this measure during each assessment period. A change of at least 5 points was estimated to be clinically meaningful [23].

### Patient-Reported Outcomes Measurement Information System: Emotional Distress-Anxiety Short Form

The questionnaire is a 4-item assessment of self-reported fear (fearfulness and panic), anxious misery (worry and dread), hyperarousal (tension, nervousness, and restlessness), and somatic symptoms related to arousal (racing heart and dizziness). Participants rated how often they experienced each item from 1 (*never*) to 5 (*always*) [24]. Participants completed this measure during each assessment period. A change of at least 1 point was estimated to be clinically meaningful [25].

### Patient-Reported Outcomes Measurement Information System: Ability to Participate in Social Roles and Activities Short Form

The questionnaire was a 4-item assessment of the perceived ability to perform one’s everyday social roles and activities. Higher scores represent fewer limitations (better abilities). Participants rated how often they had experienced each item from 5 (*never*) to 1 (*always*) [26]. Participants completed this measure during each assessment period. A change of at least 1 point was estimated to be clinically meaningful [25].

### Five Facet Mindfulness Questionnaire Nonjudging and Nonreactivity Subscales

The *Five Facet Mindfulness Questionnaire* is a 39-item assessment that examines 5 aspects of mindfulness: observing, describing, acting with awareness, nonjudging of inner experiences, and nonreactivity to inner experiences. Only the questions related to nonjudging of inner experience and nonreactivity were administered (15 items total), and participants rated whether each item was generally true for them from 1 (*never true*) to 5 (*always true*) [27]. Participants completed this measure during each assessment period. To our knowledge, estimates of minimum clinically important differences are not available.

### *Adverse Events Questions*

This questionnaire was administered in all study sessions to assess for possible adverse events and whether they were related to the study procedures (ie, “Have you experienced a negative change in your health since participating in this study?” “Have you experienced any of the following: a life-threatening event or hospitalization, or a persistent significant disruption in your ability to conduct normal life?” and “Do you think that this event was related to or caused by your participation in this study?”).

### Ethics Approval

This study was registered at ClinicalTrials.gov (NCT03844321) and was approved by the Genetic Alliance Institutional Review Board (protocol #HMHY002). Informed consent was approved by the Genetic Alliance Institutional Review Board and completed on the web by the participants.

### Statistical Analysis

We used the intention-to-treat principle for all primary analyses and incorporated all available longitudinal outcomes into the mixed effects models. We used prespecified linear mixed effects models fit via maximum likelihood to examine the comparative effectiveness of the 2 mindfulness interventions on the primary well-being endpoint (World Health Organization Five Well-Being Index [WHO-5] score): random participant slopes and intercepts; and fixed effects for intervention, time, and an intervention-by-time interaction. Our primary group comparison was based on the intervention-by-time interaction, which corresponds to the between-group difference in the slopes of average well-being scores over time. We report model-based point estimates, CIs, and *P* values. We fitted separate models to assess the intervention effects over the 8-week *intervention period* and the entire 20-week study period. Post hoc sensitivity analyses were carried out by including a fixed categorical (rather than linear) effect for time in our models, which allowed for the mean WHO-5 scores to vary over time in an unspecified fashion. We also fit post hoc marginal models via generalized estimating equations (GEEs) for the primary WHO-5 outcome, as these visually fit the raw mean trajectories more closely; these marginal GEEs used a working independence correlation structure and, similar to the prespecified mixed models, included fixed effects for intervention, time, and an intervention-by-time interaction.

Demographics and assessment scores measured at the beginning of the intervention period were analyzed as potential moderators of the relationship between treatment and well-being. The a priori moderators included (1) age; (2) role (PPRN member vs caregiver or family member); and (3) levels of well-being, stress, quality of life, anxiety, depression, and mindfulness. Several exploratory moderators were also analyzed: (1) sex, (2) ethnicity, (3) race, (4) education, (5) percentage of intervention sessions completed, (6) presence of medical problems, and (7) presence of psychiatric illness. To assess each moderator, we added fixed effects for the baseline moderator as well as moderator-by-time, moderator-by-intervention, and moderator-by-intervention-by-time interactions to the mixed model and used a likelihood ratio test to assess the 3-way interaction term, a term that represents the estimated differential intervention effect across levels of the moderator. Continuous moderators (eg, age) were assumed to have a linear moderating relationship in the moderator regression models. Moderator analyses were also carried out across both the 8-week *intervention period* and the entire 20-week study period. Statistical analyses were performed using R software (version 4.0.2, r-project.org). We used 2-sided significance levels of .05 and reported 95% CIs for all analyses.

Owing to study start-up delays, we reduced our recruitment target from 8500 to 2117 participants. We conducted separate power calculations for the original target sample size of 8500 and for the revised sample size of 2117 with PASS 14 for Cronbach α of .05 and a range of standardized mean differences (SMDs) between our MBCT and brief mindfulness groups. We chose SMDs to be consistent with what was reported by Hoffman et al [28] and allowed for some reduction in the SMDs because we assumed that the brief mindfulness group would likely receive some benefit beyond the placebo effect. For the WHO-5 (primary study endpoint), a 10-point increase on the 0- to 100-point scale (10% increase) is considered clinically significant [20,21]. On the basis of the available clinical trials using the WHO-5 (total n=3864) [28-33], the desired minimum 10-point clinically meaningful difference between the 2 treatment arms translates into SMDs ranging from 0.42 to 1.09 based on the observed SDs in the available clinical trials (SD range: 2.3-6.0) [28-33]. From this perspective, it is desirable to have a >80% statistical power to detect an SMD of 0.4.

According to our power calculations, with 8500 participants, for a Cronbach α of .05, the power would be >80% to detect SMDs >0.12% and >90% for SMDs >0.14. After revising our target sample size to 2117, for Cronbach α of .05, the power would be >80% for SMDs greater than 0.26% and >90% for SMDs greater than 0.29. In summary, with our target sample size of 2117 participants, we were able to detect differences that were even smaller than the published estimates for what constitutes a clinically meaningful change in the WHO-5.

## Results

### Study Flow and Adverse Events

Recruitment for the study occurred from February 27, 2019, to October 1, 2019, and the follow-up period ended on January 31, 2020. A total of 5029 participants were consented and completed the enrollment process. Among these, 593 (n=5029, 11.79%) individuals declined to participate, and 4436 (n=5029, 88.21%) were randomized to the study interventions. Table 1 provides the demographics of the randomized participants. A total of 25 participants requested to be removed from the study or withdrew from the study; thus, they were excluded from the data set and analyses entirely, leaving 4411 randomized participants. At week 8, a total of 496 (n=2220, 22.34%) participants in the MBCT group and 396 (n=2191, 18.07%) of participants in the brief mindfulness group completed the main outcome assessment (WHO-5). At week 20, a total of 321 (n=2220, 14.46%) participants in the MBCT group and 294 (n=2191, 13.42%) in the brief mindfulness group completed WHO-5. Completion of intervention sessions gradually decreased by session for participants in the MBCT group (week 0 completion: n=838, 38%; week 7 completion: n=317, 14%) as well as in the brief mindfulness group (week 0 completion: n=778, 36%; week 2 completion: n=418, 19%). Figure 1 shows a detailed participant flow diagram.

A total of 30 (n=2220, 1.35%) participants in the MBCT group and 31 (n=2191, 1.41%) in the brief mindfulness group reported experiencing one or more serious adverse events during the full study period. In addition, 33 (n=2220, 1.49%) participants in the MBCT group and 41 (n=2191, 1.87%) in the brief mindfulness group reported experiencing one or more *nonserious* adverse events. No serious adverse events or nonserious adverse events were reported by participants to be related to the study. The most common category of events reported was a negative life event that was unrelated to the study (eg, death of family members; n=23 participants).

**Table 1.** Demographics of randomized participants.

| Characteristic[a] | Brief mindfulness (n=2191) | MBCT[b] (n=2220) | Total (n=4411) |
|---|---|---|---|
| Age (years), mean (SD) | 54.11 (15.05) | 55.44 (14.72) | 54.78 (14.90) |
| **Race or origin, n (%)** | | | |
| White | 1916 (87.45) | 1957 (88.15) | 3873 (87.80) |
| Multiple race[c] | 59 (2.69) | 68 (3.06) | 127 (2.88) |
| Black, African American, African, or Afro-Caribbean | 52 (2.37) | 45 (2.03) | 97 (2.20) |
| Asian | 38 (1.73) | 39 (1.76) | 77 (1.75) |
| Other | 24 (1.10) | 32 (1.44) | 56 (1.27) |
| Native American, American Indian, or Alaska Native | 15 (0.68) | 12 (0.54) | 27 (0.61) |
| Unknown | 2 (0.09) | 5 (0.23) | 7 (0.16) |
| Native Hawaiian or other Pacific Islander | 1 (0.05) | 2 (0.09) | 3 (0.07) |
| No selected answer | 41 (1.87) | 36 (1.62) | 77 (1.75) |
| Prefer not to answer | 43 (1.96) | 24 (1.08) | 67 (1.52) |
| **Sex assigned at birth, n (%)** | | | |
| Female | 1767 (80.65) | 1780 (80.18) | 3547 (80.41) |
| Male | 410 (18.71) | 429 (19.32) | 839 (19.02) |
| Unknown | 8 (0.37) | 5 (0.23) | 13 (0.29) |
| Ambiguous | 5 (0.23) | 6 (0.27) | 11 (0.25) |
| Other | 1 (0.05) | 0 (0.00) | 1 (0.02) |
| **Sexual orientation, n (%)** | | | |
| Straight | 1743 (79.55) | 1816 (81.80) | 3559 (80.68) |
| Bisexual | 102 (4.66) | 91 (4.10) | 193 (4.38) |
| Gay | 78 (3.56) | 72 (3.24) | 150 (3.40) |
| Lesbian | 78 (3.56) | 69 (3.11) | 147 (3.33) |
| Queer | 67 (3.06) | 56 (2.52) | 123 (2.79) |
| Asexual | 22 (1.00) | 18 (0.81) | 40 (0.91) |
| Multiple sexual orientations | 8 (0.37) | 14 (0.63) | 22 (0.50) |
| Something else | 14 (0.64) | 5 (0.23) | 19 (0.43) |
| Questioning | 7 (0.32) | 9 (0.41) | 16 (0.36) |
| Other | 6 (0.27) | 3 (0.14) | 9 (0.20) |
| Unknown | 1 (0.05) | 4 (0.18) | 5 (0.11) |
| No selected answer | 35 (1.60) | 39 (1.76) | 74 (1.68) |
| Prefer not to answer | 30 (1.37) | 24 (1.08) | 54 (1.22) |
| Total sexual minority[d] | 382 (17.43) | 337 (15.18) | 719 (16.30) |
| **Gender identity, n (%)** | | | |
| Woman | 1653 (75.45) | 1690 (76.13) | 3343 (75.79) |
| Man | 388 (17.71) | 402 (18.11) | 790 (17.91) |
| Genderqueer | 39 (1.78) | 32 (1.44) | 71 (1.61) |
| Transgender male, trans man, or female-to-male | 32 (1.46) | 17 (0.77) | 49 (1.11) |
| Other | 8 (0.37) | 8 (0.36) | 16 (0.36) |
| Something else | 8 (0.37) | 4 (0.18) | 12 (0.27) |
| Transgender female, trans women, or male-to-female | 4 (0.18) | 6 (0.27) | 10 (0.23) |

| Characteristic[a] | Brief mindfulness (n=2191) | MBCT[b] (n=2220) | Total (n=4411) |
|---|---|---|---|
| Multiple gender categories | 4 (0.18) | 5 (0.23) | 9 (0.20) |
| Unknown | 2 (0.09) | 1 (0.05) | 3 (0.07) |
| No selected answer | 43 (1.96) | 49 (2.21) | 92 (2.09) |
| Prefer not to answer | 10 (0.46) | 6 (0.27) | 16 (0.36) |
| **Ethnicity, n (%)** | | | |
| Non-Hispanic | 2041 (93.15) | 2078 (93.60) | 4119 (93.38) |
| Hispanic | 97 (4.43) | 101 (4.55) | 198 (4.49) |
| Other | 5 (0.23) | 2 (0.09) | 7 (0.16) |
| Unknown | 2 (0.09) | 3 (0.14) | 5 (0.11) |
| No selected answer | 22 (1.00) | 20 (0.90) | 42 (0.95) |
| Prefer not to answer | 24 (1.10) | 16 (0.72) | 40 (0.91) |
| **Role, n (%)** | | | |
| Patient-Powered Research Network member | 2066 (94.29) | 2087 (94.01) | 4153 (94.15) |
| Family member or caregiver | 125 (5.71) | 133 (5.99) | 258 (5.85) |

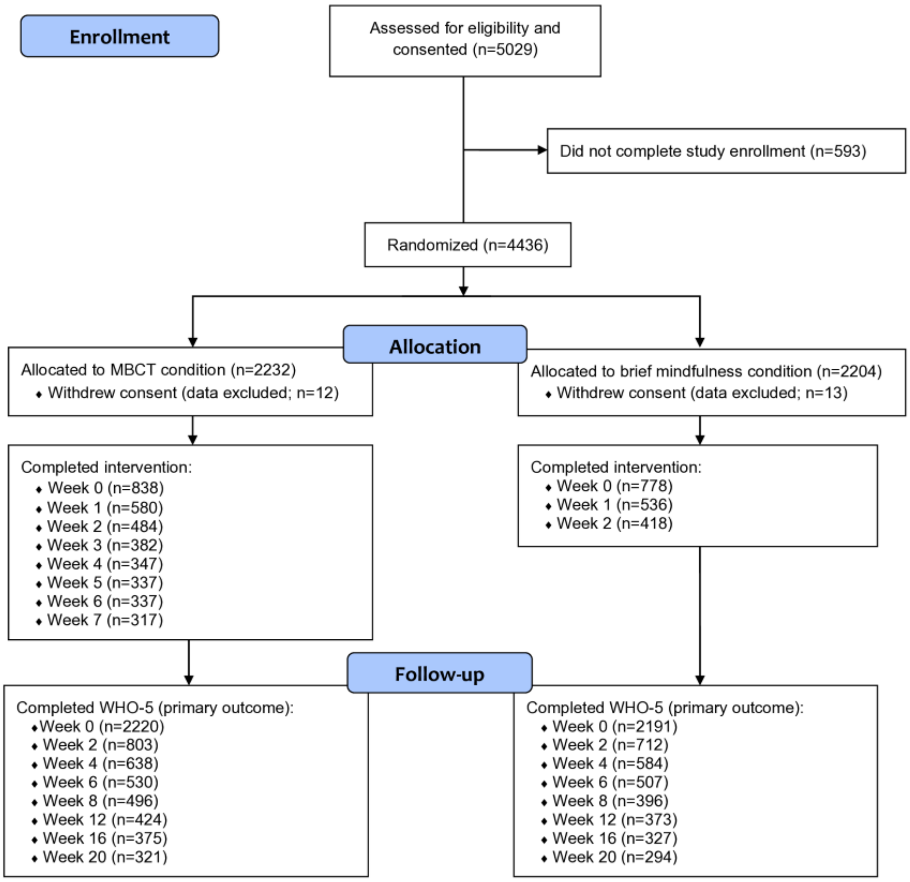
[a] Participants could select one option for each demographic question.

[b] MBCT: mindfulness-based cognitive therapy.

[c] Two or more races.

[d] Bisexual, gay, lesbian, queer, asexual, multiple sexual orientations, something else, questioning, or other.

**Figure 1.** Enrollment flow diagram. An intervention session was marked as complete only if the participant had completed the entire session. Participants were not allowed to skip any page during the intervention sessions. MBCT: mindfulness-based cognitive therapy; WHO-5: World Health Organization—Five Well-Being Index.



## Primary Outcome: Well-being

Figure 2 shows a graph of the average well-being scores by condition across the entire study period. The average well-being scores improved for both the MBCT group and the brief mindfulness intervention group over the 8-week *intervention period* and entire 20-week study periods (Tables 2 and 3). For example, based on the 20-week model, average WHO-5 scores increased by 0.41 (95% CI 0.34 to 0.48) points per week in the MBCT group and 0.33 (95% CI 0.26 to 0.40) points per week in the brief mindfulness group. Changes in well-being were not significantly different for the 8-session MBCT group compared with the brief 3-session mindfulness group over either the 8- or 20-week period (*P*=.80 and .10, respectively; Tables 2 and 3). Similarly, no differences between groups over time were found when allowing for nonlinear trajectories in mean WHO-5 scores in mixed models (*P*=.47 and .16, respectively) or when using a marginal model fit via GEE with linear time (*P*=.78 and .77, respectively) or categorical time (*P*=.51 and .83, respectively; Figure S1; Multimedia Appendix 1, Table S1).

Regarding potential moderators, only age and percentage of intervention sessions over the intervention period suggested differential changes in well-being scores between the 2 conditions (Multimedia Appendix 2). For each continuous moderator, we reported model-based estimated changes in well-being by intervention group (and between-group differences) at values corresponding to the 25th, 50th, and 75th percentiles of each moderator at baseline (recall, continuous moderators were assumed to have a linear relationship with the differential effect of treatment). Specifically, the estimated differential improvement in well-being comparing MBCT with the brief mindfulness program over 8 weeks was more pronounced in younger people (*P*=.05) and those with a higher percentage of intervention sessions (*P*=.005); these differential effects were not sustained over the full 20-week study period. For all other moderators considered (sex assigned at birth, gender, sexual orientation, ethnicity, race, education, baseline perceived stress, baseline depression, baseline anxiety, baseline perceived ability to perform social roles, baseline mindfulness, presence of medical problems, and presence of psychiatric illness), there was no evidence of a differential effect over either the 8- or 20-week period (Multimedia Appendix 2).

**Figure 2.** Well-being by week by intervention group. The total WHO-5 score can range from 0 to 100, although we restricted the y-axis to the 25% and 75% quantiles (40 and 72, respectively) of the WHO-5 scores reported in this study. Points correspond to sample means and vertical lines correspond to pointwise 95% CIs for the means. MBCT: mindfulness-based cognitive therapy; WHO-5: World Health Organization—Five Well-Being Index.

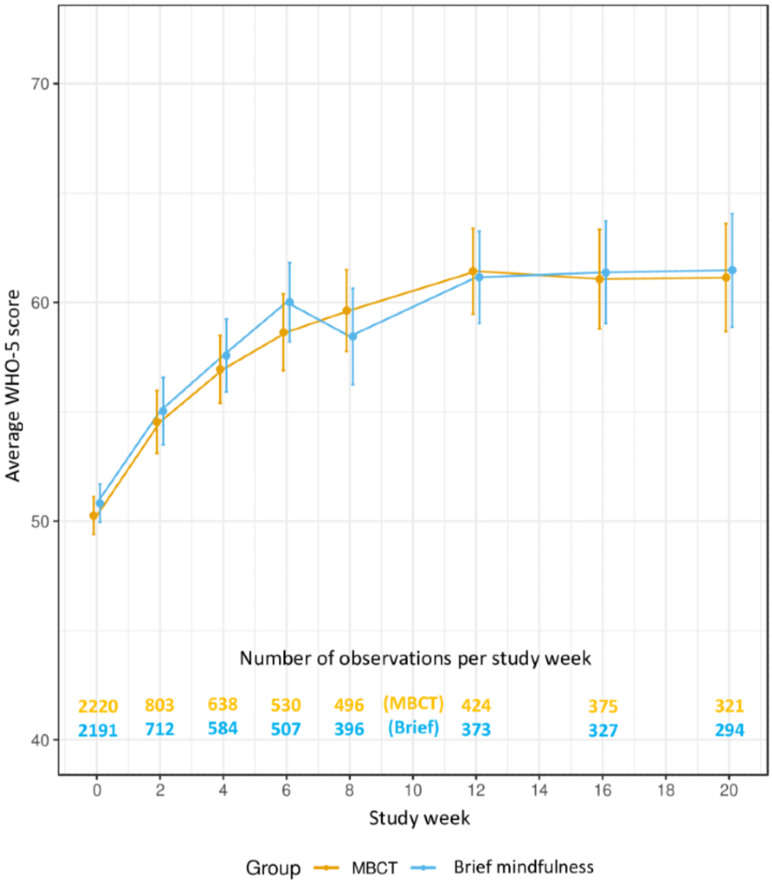


**Table 2.** Mixed model–based change in mean outcome per week by treatment group and period.

| Outcome | Brief mindfulness slope estimate (95% CI) | Mindfulness-based cognitive therapy slope estimate (95% CI) | Difference in slopes | |
|---|---|---|---|---|
| | | | Estimate (95% CI) | *P* value |
| **Intervention period (baseline to 8 weeks)** | | | | |
| Well-being | 0.76 (0.60 to 0.91) | 0.78 (0.63 to 0.93) | −0.02 (−0.24 to 0.19) | .80 |
| Anxiety | −0.13 (−0.16 to −0.11) | −0.12 (−0.15 to −0.10) | −0.01 (−0.05 to 0.02) | .40 |
| Depression | −0.16 (−0.21 to −0.12) | −0.19 (−0.23 to −0.15) | 0.02 (−0.04 to 0.09) | .43 |
| Perceived ability to perform social roles | 0.14 (0.11 to 0.17) | 0.12 (0.09 to 0.15) | 0.01 (−0.03 to 0.06) | .50 |
| Perceived stress | −0.15 (−0.18 to −0.11) | −0.13 (−0.16 to −0.09) | −0.02 (−0.07 to 0.02) | .34 |
| FFMQ[a]: nonjudging | 0.48 (0.42 to 0.53) | 0.53 (0.48 to 0.59) | −0.06 (−0.14 to 0.02) | .13 |
| FFMQ: nonreactivity | 0.41 (0.36 to 0.46) | 0.40 (0.35 to 0.44) | | .63 |
| **Study period (baseline to 20 weeks)** | | | | |
| Well-being | 0.33 (0.26 to 0.40) | 0.41 (0.34 to 0.48) | −0.08 (−0.18 to 0.02) | .10 |
| Anxiety | −0.05 (−0.06 to −0.04) | −0.07 (−0.08 to −0.05) | 0.02 (0.0 to 0.03) | .05 |
| Depression | −0.06 (−0.08 to −0.04) | −0.11 (−0.13 to −0.09) | 0.05 (0.02 to 0.08) | <.001 |
| Perceived ability to perform social roles | 0.04 (0.03 to 0.05) | 0.05 (0.04 to 0.06) | −0.01 (0.03 to 0.01) | .43 |
| Perceived stress | −0.07 (−0.08 to −0.05) | −0.08 (−0.09 to −0.06) | 0.01 (−0.01 to 0.03) | .35 |
| FFMQ: nonjudging | 0.20 (0.18 to 0.23) | 0.24 (0.21 to 0.26) | −0.04 (−0.07 to −0.0) | .04 |
| FFMQ: nonreactivity | 0.18 (0.15 to 0.20) | 0.20 (0.18 to 0.22) | −0.02 (−0.05 to 0.01) | .20 |

[a] FFMQ: Five Facet Mindfulness Questionnaire.

**Table 3.** Mixed model–based change in mean outcome from baseline by treatment group and period.

| Outcome | Brief mindfulness change from baseline estimate (95% CI) | MBCT[a] change from baseline estimate (95% CI) | Difference in changes from baseline (brief—MBCT) estimate (95% CI) |
|---|---|---|---|
| **Intervention period (baseline to 8 weeks)** | | | |
| Well-being | 6.03 (4.8 to 7.26) | 6.25 (5.07 to 7.42) | −0.22 (−1.92 to 1.48) |
| Anxiety | −1.07 (−1.27 to −0.87) | −0.95 (−1.14 to −0.76) | −0.12 (−0.4 to 0.16) |
| Depression | −1.32 (−1.68 to −0.96) | −1.52 (−1.86 to −1.17) | 0.2 (−0.3 to 0.7) |
| Perceived ability to perform social roles | 1.1 (0.86 to 1.35) | 0.99 (0.75 to 1.22) | 0.12 (−0.22 to 0.46) |
| Perceived stress | −1.18 (−1.45 to −0.91) | −1 (−1.25 to −0.75) | −0.18 (−0.55 to 0.19) |
| FFMQ[b]: nonjudging | 3.8 (3.36 to 4.24) | 4.27 (3.85 to 4.69) | −0.47 (−1.08 to 0.14) |
| FFMQ: nonreacting | 3.3 (2.92 to 3.68) | 3.17 (2.81 to 3.53) | 0.13 (−0.4 to 0.65) |
| **Study period (baseline to 20 weeks)** | | | |
| Well-being | 6.55 (5.1 to 8) | 8.22 (6.84 to 9.6) | −1.67 (−3.67 to 0.33) |
| Anxiety | −0.99 (−1.22 to −0.76) | −1.32 (−1.54 to −1.1) | 0.32 (0.01 to 0.64) |
| Depression | −1.12 (−1.54 to −0.7) | −2.16 (−2.56 to −1.76) | 1.04 (0.46 to 1.62) |
| Perceived ability to perform social roles | 0.82 (0.54 to 1.09) | 0.97 (0.71 to 1.23) | −0.15 (−0.53 to 0.22) |
| Perceived stress | −1.32 (−1.61 to −1.04) | −1.52 (−1.79 to −1.24) | 0.19 (−0.21 to 0.59) |
| FFMQ: nonjudging | 4.03 (3.55 to 4.52) | 4.75 (4.29 to 5.21) | −0.72 (−1.39 to −0.04) |
| FFMQ: nonreacting | 3.52 (3.1 to 3.94) | 3.9 (3.5 to 4.3) | −0.38 (−0.96 to 0.2) |

[a] MBCT: mindfulness-based cognitive therapy.

[b] FFMQ: Five Facet Mindfulness Questionnaire.

## Secondary Outcomes

For both treatment conditions, all secondary outcomes of anxiety, depression, perceived ability to perform social roles, perceived stress, and mindfulness improved over the intervention and study periods (Tables 2 and 3). Although we found no between-group differences in secondary outcomes over 8 weeks, the average improvements in depression (*P*<.001), anxiety (*P*=.05), and mindfulness (*P*=.03) were greater in the MBCT group than in the brief mindfulness intervention group over the full 20-week study period.

## Predictors of Dropout

Participants who completed the study, defined for these analyses as randomized participants who provided at least one WHO-5 score at a postintervention follow-up visit (ie, among visit weeks 12, 16, or 20), significantly differed from those who did not complete the study on several baseline and clinical characteristics. Completers were disproportionately assigned male sex at birth (*P*<.001), older (*P*<.001), straight (*P*<.001), more highly educated (*P*=.04), and tended to have lower baseline depression (*P*<.001), anxiety (*P*<.001), and well-being (*P*<.001) and higher perceived ability to perform social roles (*P*<.001) and mindfulness (*P*<.001) than those who did not complete the study (Table 4).

**Table 4.** Baseline demographics and clinical characteristics of dropouts versus completers.

| | Dropouts (N=3300) | Completers[a] (N=1111) | Overall, (N=4411) | *P* value[b] |
|---|---|---|---|---|
| **Sex assigned at birth, n (%)** | | | | <.001 |
| Male | 579 (17.55) | 260 (23.4) | 839 (19.02) | |
| Female | 2705 (81.97) | 842 (75.79) | 3547 (80.41) | |
| Other, unknown, or ambiguous | 16 (0.48) | 9 (0.81) | 25 (0.57) | |
| **Sexual orientation, n (%)** | | | | <.001 |
| Straight | 2623 (79.48) | 936 (84.25) | 3559 (80.68) | |
| Sexual minority | 677 (20.52) | 175 (15.75) | 852 (19.32) | |
| **Gender identity, n (%)** | | | | .24 |
| Cisgender | 3068 (92.97) | 1045 (94.06) | 4113 (93.24) | |
| Gender minority | 232 (7.03) | 66 (5.94) | 298 (6.76) | |
| **Randomization, n (%)** | | | | .18 |
| MBCT[c] | 1641 (49.73) | 579 (52.12) | 2220 (50.33) | |
| Brief mindfulness | 1659 (50.27) | 532 (47.88) | 2191 (49.67) | |
| **Hispanic, n (%)** | | | | .24 |
| Yes | 158 (4.84) | 40 (3.63) | 198 (4.53) | |
| No | 3069 (93.97) | 1050 (95.19) | 4119 (94.28) | |
| Other, unknown, or prefer not to answer | 39 (1.19) | 13 (1.18) | 52 (1.19) | |
| **Race, n (%)** | | | | .06 |
| White | 2876 (88.85) | 997 (90.88) | 3873 (89.36) | |
| Other[d] | 361 (11.15) | 100 (9.12) | 461 (10.64) | |
| **School, n (%)** | | | | .04 |
| High school or less | 184 (5.71) | 44 (4.03) | 228 (5.29) | |
| 2- or 4-year college | 1627 (50.53) | 537 (49.13) | 2164 (50.17) | |
| More than 4-year college | 1409 (43.76) | 512 (46.84) | 1921 (44.54) | |
| Age (years), mean (SD) | 53.42 (14.95) | 58.83 (13.98) | 54.78 (14.90) | <.001 |
| Depression, mean (SD) | 16.82 (7.12) | 15.06 (6.62) | 16.38 (7.04) | <.001 |
| Anxiety, mean (SD) | 9.18 (3.67) | 8.23 (3.44) | 8.94 (3.64) | <.001 |
| Perceived ability to perform social roles, mean (SD) | 13.24 (4.29) | 14.14 (4.24) | 13.47 (4.29) | <.001 |
| Well-being, mean (SD) | 49.11 (20.51) | 54.76 (20.80) | 50.53 (20.73) | <.001 |
| FFMQ[e]: nonjudging, mean (SD) | 28.16 (7.48) | 29.64 (7.34) | 28.54 (7.47) | <.001 |
| FFMQ: nonreacting, mean (SD) | 21.07 (5.64) | 22.17 (5.66) | 21.35 (5.67) | <.001 |

[a] Completers were defined as randomized participants who provided at least one WHO-5 score at a postintervention follow-up visit (ie, among visit weeks 12, 16, or 20).

[b] *P* value based on Fisher exact test or unequal variances *t* test, as appropriate.

[c] MBCT: mindfulness-based cognitive therapy.

[d] Others included (1) native American, American Indian, or Alaskan Native; (2) Asian; (3) Black, African American, African, or Afro-Caribbean; (4) Native Hawaiian or other Pacific Islander; (5) multiple race; (6) other; (7) unknown; (8) prefer not to answer.

[e] FFMQ: Five Facet Mindfulness Questionnaire.

## Discussion

### Principal Findings

This study demonstrated that although both a standard 8-session MBCT program and a shorter 3-session mindfulness program mildly improved overall well-being scores over the 8- and 20-week periods, it did not support our hypothesis that the standard MBCT program would yield superior results with regard to the primary outcome of well-being. Participants in the MBCT program experienced statistically greater improvements in depression, anxiety, and mindfulness compared with their brief mindfulness counterparts, but these group differences are unlikely to be clinically meaningful [34-38].

Younger people and participants who completed a higher proportion of intervention sessions reported larger improvements in overall well-being, an effect that was more pronounced for participants assigned to the standard MBCT intervention. Although the standard MBCT did not prove superior to brief mindfulness in improving participant well-being in aggregate, these findings suggest that it could be a better choice for younger people as well as treatment-adherent individuals.

Among the secondary outcomes, participants in the standard MBCT condition had significantly greater improvements in anxiety, depression, and mindfulness scores than those in the brief mindfulness condition. This may be evidence for the superiority of standard MBCT in treating anxiety and depression as well as in improving mindfulness. However, these results should be interpreted with caution, given that the group differences were minimal and unlikely to be clinically meaningful.

## Comparison With Prior Work

These findings are consistent with a study comparing the efficacy of an 8-week in-person standard and a 4-week in-person abbreviated mindfulness-based intervention in adult undergraduate students (N=99) that found no significant differences in improvements to participant mindfulness, as well as self-compassion, positive and negative affect, anxiety and depressive symptoms, and resilience [39]. Another study found that a brief version of an in-person MBSR intervention demonstrated equally significant decreases in perceived stress and improvements in sleep quality compared with a traditional MBSR intervention in a study of 48 healthy adults [40]. In addition, our findings that participants with more severe mental health symptoms at baseline (ie, more severe depression and anxiety) and lower baseline well-being and mindfulness were more likely to drop out of the study are consistent with the findings of several studies that greater baseline symptom severity may predict attrition in web-based interventions [41-43].

Although our study did not support the hypothesis that the longer standard 8-session MBCT would be superior to the shorter mindfulness intervention, it is possible that a shorter mindfulness program is sufficient for improving well-being across a variety of patient populations [14,44]. For example, patients with cardiac disease (N=114) were randomized to a 4-session in-person mindfulness group or to a self-help control group and yielded better outcomes for the mindfulness group on quality of life, anxiety, depression, and perceived stress, which were partially or fully mediated by an increase in mindfulness [14]. A shorter, 2-session mindfulness group to reduce alcohol consumption was also found to be effective among college students [44]. In a study that assessed the efficacy of a web-based MBCT course [45], participants (n=118) who completed the course demonstrated significant improvements in perceived stress, depression, and anxiety, which were maintained at the 3- and 6-month follow-ups [46]. Given the wide range of conditions and special populations represented in our study sample, our results suggest that shorter mindfulness programs could yield important benefits across the general population of individuals with various medical conditions. The similar benefits experienced by the participants in the 2 treatment conditions may be because of effects other than the intended mindfulness therapy. However, this is unlikely, as prior research supports that both interventions adapted for this study are superior to placebo [8,11].

The strengths of this study include leveraging existing registries of individuals. This allowed us to recruit and consent 5029 participants quickly (ie, over 8 months), with a mean of 625 participants per month. Efficient recruitment was likely due to the broad inclusion and exclusion criteria, including stakeholder input, collaboration with PPRNs already engaged in a collaborative network, and conducting a web-based study with broad outreach.

## Limitations

The main limitation is the high attrition and low completion rates for both the intervention sessions and assessments. Although our primary results, which are based on likelihood-based mixed effects regression models, are valid under the missing at random assumption (ie, statistical independence of outcomes and missingness conditional on prior observed outcome measurements, treatment group, and time), the results could be inaccurate if endpoints are missing *not* at random. This study was also limited by a lack of diversity in the study population, especially with regard to race and gender. Our results may not be generalizable to men or to people of different colors. The second limitation was the lack of support or guidance for the participants, such as technical support with the study platform or assistance with the intervention material. Third, all outcomes were self-reported, and we did not include diagnostic assessments (which could have better characterized the sample) to minimize participant burden. Fourth, given the large sample size, we were not able to reimburse the participants' time in completing the assessments (with the exception of a raffle) or in completing the web-based intervention. The lack of incentives and technical support for the participants could have contributed to the high attrition rates. However, these limitations increase the generalizability and applicability to more general populations.

## Conclusions

In summary, this study demonstrated that a web-based 8-session MBCT program was not superior to a web-based brief 3-session mindfulness program in improving the well-being of participants. Younger people and treatment-adherent individuals (ie, those who attend a higher percentage of sessions) may benefit more from standard-length MBCT.

## Acknowledgments

The authors would like to thank the participants of the Healthy Mind Healthy You study for sharing their valuable time, information, and experience. The authors would also like to thank their collaborators at the Patient-Centered Research Outcomes Institute Patient-Powered Research Networks (PPRN), as well as their other stakeholders in the Healthy Mind Healthy You study for their help in all aspects of the study as well as for their input into this manuscript. Finally, the authors would like to thank Nora Horick, Daniel Nolte, and the biostatistical team at Massachusetts General Hospital. The authors are grateful for the collaboration. This research was funded by the Patient-Centered Outcomes Research Institute Program Award XPPRN-1512-33786. The study and full study protocol can be found at ClinicalTrials.gov (title: Healthy Mind Healthy You: A Study of Mindfulness; ClinicalTrials.gov identifier: NCT03844321).

## Conflicts of Interest

LGS receives grant funding from National Institutes of Health and Patient Centered Outcomes Research Institute as well as royalties from New Harbinger. She has also served on the Scientific Advisory Boards for Depression and Bipolar Support Alliance and the Milken Institute. AAN received funding as a consultant and scientific advisory board member for headspace health. In accordance with the Patient-Centered Research Outcomes Institute policy and federal law, portions of this manuscript will be published on Patient-Centered Research Outcomes Institute's website as a final research report. In the past year, SVF has either received income, potential income, travel expenses, continuing education support, research support or several of these aids from Aardvark, Akili, Genomind, Ironshore, KemPharm/Corium, Noven, Ondosis, Otsuka, Rhodes, Supernus, Takeda, Tris, and Vallon. With his institution, he has US patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of Attention Deficit Hyperactivity Disorder. In previous years, he received support from Alcobra, Arbor, Aveksham, CogCubed, Eli Lilly, Enzymotec, Impact, Janssen, Lundbeck/Takeda, McNeil, NeuroLifeSciences, Neurovance, Novartis, Pfizer, Shire, and Sunovion. He also receives royalties from books published by Guilford Press: "Straight Talk about Your Child's Mental Health"; Oxford University Press: "Schizophrenia: The Facts"; and Elsevier: "ADHD: Non-Pharmacologic Interventions". In addition, he is the program director of Attention Deficit Hyperactivity Disorder in Adults.

Multimedia Appendix 1
Supplemental material for publication including generalized estimating equation (GEE) model–based change in mean outcome per week by treatment group and period, participating patient-powered research networks and respective recruitment numbers, study assessment schedule and intervention sessions, model-based change in mean outcome per week by treatment group and period additionally accounting for patient-powered research network (PPRN) as a random effect, questionnaire scores at each assessment timepoint, crude standardized effect sizes by treatment group, World Health Organization—Five Well-Being Index (WHO-5) scores at each assessment timepoint by treatment group, well-being by week by intervention group with mixed and GEE model-based estimates, and screenshots of brief mindfulness program by week.
[DOCX File , 2074 KB - jmir_v24i9e35620_app1.docx ]

Multimedia Appendix 2
Heterogeneity of treatment effects.
[DOCX File , 29 KB - jmir_v24i9e35620_app2.docx ]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 966 KB - jmir_v24i9e35620_app3.pdf ]

## References

1. World happiness report. 2022. URL: https://worldhappiness.report/ [accessed 2020-08-12]
2. Witters D. Hawaii Tops U.S. in Wellbeing for Record 7th Time. Gallup. 2019 Feb 27. URL: https://news.gallup.com/poll/247034/hawaii-tops-wellbeing-record-7th-time.aspx#:~:text=WASHINGTON%2C%20D.C.%20%2D%2D%20Hawaii%20residents,two%20states%20to%20do%20so [accessed 2020-08-12]
3. Anderson NB, Johnson SB, Belar CD, Breckler SJ, Nordal KC, Ballard D, et al. Stress in America™: Our Health at Risk. American Psychological Association. 2012 Jan 11. URL: https://www.apa.org/news/press/releases/stress/2011/final-2011.pdf [accessed 2016-02-22]
4. Stulz N, Lutz W, Kopta SM, Minami T, Saunders SM. Dose-effect relationship in routine outpatient psychotherapy: does treatment duration matter? J Couns Psychol 2013 Oct;60(4):593-600. [doi: 10.1037/a0033589] [Medline: 23815633]
5. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
6. Howard KI, Kopta SM, Krause MS, Orlinsky DE. The dose-effect relationship in psychotherapy. Am Psychol 1986 Feb;41(2):159-164. [Medline: 3516036]

7. Zeidan F, Johnson SK, Diamond BJ, David Z, Goolkasian P. Mindfulness meditation improves cognition: evidence of brief mental training. Conscious Cogn 2010 Jun;19(2):597-605. [doi: 10.1016/j.concog.2010.03.014] [Medline: 20363650]
8. Banks JB, Welhaf MS, Srour A. The protective effects of brief mindfulness meditation training. Conscious Cogn 2015 May;33:277-285. [doi: 10.1016/j.concog.2015.01.016] [Medline: 25680006]
9. Creswell JD, Pacilio LE, Lindsay EK, Brown KW. Brief mindfulness meditation training alters psychological and neuroendocrine responses to social evaluative stress. Psychoneuroendocrinology 2014 Jun;44:1-12. [doi: 10.1016/j.psyneuen.2014.02.007] [Medline: 24767614]
10. Jain S, Shapiro SL, Swanick S, Roesch SC, Mills PJ, Bell I, et al. A randomized controlled trial of mindfulness meditation versus relaxation training: effects on distress, positive states of mind, rumination, and distraction. Ann Behav Med 2007 Feb;33(1):11-21. [doi: 10.1207/s15324796abm3301_2] [Medline: 17291166]
11. Cathcart S, Galatis N, Immink M, Proeve M, Petkov J. Brief mindfulness-based therapy for chronic tension-type headache: a randomized controlled pilot study. Behav Cogn Psychother 2014 Jan;42(1):1-15. [doi: 10.1017/S1352465813000234] [Medline: 23552390]
12. Kingston J, Chadwick P, Meron D, Skinner TC. A pilot randomized control trial investigating the effect of mindfulness practice on pain tolerance, psychological well-being, and physiological activity. J Psychosom Res 2007 Mar;62(3):297-300. [doi: 10.1016/j.jpsychores.2006.10.007] [Medline: 17324679]
13. Cruess DG, Finitsis DJ, Smith A, Goshe BM, Burnham K, Burbridge C, et al. Brief stress management reduces acute distress and buffers physiological response to a social stress test. Int J Stress Man 2015;22(3):270-286. [doi: 10.1037/a0039130]
14. Nyklíček I, Dijksman SC, Lenders PJ, Fonteijn WA, Koolen JJ. A brief mindfulness based intervention for increase in emotional well-being and quality of life in percutaneous coronary intervention (PCI) patients: the MindfulHeart randomized controlled trial. J Behav Med 2014 Feb;37(1):135-144. [doi: 10.1007/s10865-012-9475-4] [Medline: 23180285]
15. Carman KL. A New Path for PCORnet's Patient-Powered Research Networks. Patient-Centered Outcomes Research Institute. 2018 Oct 22. URL: https://www.pcori.org/blog/new-path-pcornets-patient-powered-research-networks [accessed 2020-08-20]
16. Sylvia LG, Hearing CM, Montana RE, Gold AK, Walsh SL, Janos JA, et al. MoodNetwork: an innovative approach to patient-centered research. Med Care 2018 Oct;56 Suppl 10 Suppl 1:S48-S52 [FREE Full text] [doi: 10.1097/MLR.0000000000000789] [Medline: 30074951]
17. Segal ZV, Williams JM, Teasdale JD. Mindfulness-Based Cognitive Therapy for Depression. 2nd edition. New York, NY, USA: The Guilford Press; 2013.
18. Zeidan F, Gordon NS, Merchant J, Goolkasian P. The effects of brief mindfulness meditation training on experimentally induced pain. J Pain 2010 Mar;11(3):199-209. [doi: 10.1016/j.jpain.2009.07.015] [Medline: 19853530]
19. Deckersbach TH, Hölzel B, Eisner L, Lazar SW, Nierenberg AA. Mindfulness-Based Cognitive Therapy for Bipolar Disorder. New York, NY, USA: The Guildford Press; 2014.
20. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
21. Awata S, Bech P, Yoshida S, Hirai M, Suzuki S, Yamashita M, et al. Reliability and validity of the Japanese version of the World Health Organization-Five Well-Being Index in the context of detecting depression in diabetic patients. Psychiatry Clin Neurosci 2007 Feb;61(1):112-119 [FREE Full text] [doi: 10.1111/j.1440-1819.2007.01619.x] [Medline: 17239048]
22. LEVEL 2–Depression–Adult (PROMIS emotional distress–depression–Short form). PROMIS Health Organization (PHO) and PROMIS Cooperative Group. Washington, DC, USA: American Psychiatric Association; 2013. URL: https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM5_Level-2-Depression-Adult.pdf [accessed 2016-02-23]
23. Beleckas CM, Guattery J, Chamberlain AM, Khan T, Kelly MP, Calfee RP. Using patient-reported outcomes measurement information system measures to understand the relationship between improvement in physical function and depressive symptoms. J Am Acad Orthop Surg 2018 Dec 15;26(24):e511-e518 [FREE Full text] [doi: 10.5435/JAAOS-D-17-00039] [Medline: 30216242]
24. LEVEL 2—Anxiety—Adult (PROMIS Emotional Distress—Anxiety— Short Form). PROMIS Health Organization (PHO) and PROMIS Cooperative Group. Washington, DC, USA: American Psychiatric Association; 2013. URL: https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/DSM-5-TR/APA-DSM5TR-Level2AnxietyAdult.pdf [accessed 2016-02-23]
25. Chen CX, Kroenke K, Stump TE, Kean J, Carpenter JS, Krebs EE, et al. Estimating minimally important differences for the PROMIS pain interference scales: results from 3 randomized clinical trials. Pain 2018 Apr;159(4):775-782 [FREE Full text] [doi: 10.1097/j.pain.0000000000001121] [Medline: 29200181]
26. LEVEL 2 — Depression — Adult (PROMIS: Ability to Participate in Social Roles and Activities-Short Form). PROMIS Health Organization and PROMIS Cooperative Group. 2016. URL: https://www.apaservices.org/practice/reimbursement/health-registry/adult-ability-participate-social.pdf [accessed 2016-02-23]
27. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]

28. Hoffman CJ, Ersser SJ, Hopkinson JB, Nicholls PG, Harrington JE, Thomas PW. Effectiveness of mindfulness-based stress reduction in mood, breast- and endocrine-related quality of life, and well-being in stage 0 to III breast cancer: a randomized, controlled trial. J Clin Oncol 2012 Apr 20;30(12):1335-1342. [doi: 10.1200/JCO.2010.34.0331] [Medline: 22430268]
29. Guico-Pabia CJ, Fayyad RS, Soares CN. Assessing the relationship between functional impairment/recovery and depression severity: a pooled analysis. Int Clin Psychopharmacol 2012 Jan;27(1):1-7. [doi: 10.1097/YIC.0b013e32834c2488] [Medline: 21971533]
30. Logtenberg SJ, Kleefstra N, Houweling ST, Groenier KH, Gans RO, Bilo HJ. Health-related quality of life, treatment satisfaction, and costs associated with intraperitoneal versus subcutaneous insulin administration in type 1 diabetes: a randomized controlled trial. Diabetes Care 2010 Jun;33(6):1169-1172 [FREE Full text] [doi: 10.2337/dc09-1758] [Medline: 20185731]
31. Martiny K, Refsgaard E, Lund V, Lunde M, Sørensen L, Thougaard B, et al. A 9-week randomized trial comparing a chronotherapeutic intervention (wake and light therapy) to exercise in major depressive disorder patients treated with duloxetine. J Clin Psychiatry 2012 Sep;73(9):1234-1242. [doi: 10.4088/JCP.11m07625] [Medline: 23059149]
32. Robinson SK, Viirre ES, Bailey KA, Gerke MA, Harris JP, Stein MB. Randomized placebo-controlled trial of a selective serotonin reuptake inhibitor in the treatment of nondepressed tinnitus subjects. Psychosom Med 2005;67(6):981-988. [doi: 10.1097/01.psy.0000188479.04891.74] [Medline: 16314604]
33. Wade AG, Ford I, Crawford G, McMahon AD, Nir T, Laudon M, et al. Efficacy of prolonged release melatonin in insomnia patients aged 55-80 years: quality of sleep and next-day alertness outcomes. Curr Med Res Opin 2007 Oct;23(10):2597-2605. [doi: 10.1185/030079907X233098] [Medline: 17875243]
34. Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. Responsiveness and minimally important differences for 4 patient-reported outcomes measurement information system short forms: physical function, pain interference, depression, and anxiety in knee osteoarthritis. J Pain 2017 Sep;18(9):1096-1110 [FREE Full text] [doi: 10.1016/j.jpain.2017.05.001] [Medline: 28501708]
35. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol 2011 May;64(5):507-516 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.11.018] [Medline: 21447427]
36. Swanholm E, McDonald W, Makris U, Noe C, Gatchel R. Estimates of minimally important differences (MIDs) for two patient-reported outcomes measurement information system (PROMIS) computer-adaptive tests in chronic pain patients. J Appl Biobehav Res 2014 Dec 03;19(4):217-232 [FREE Full text] [doi: 10.1111/jabr.12026]
37. Kroenke K, Baye F, Lourens SG. Comparative responsiveness and minimally important difference of common anxiety measures. Med Care 2019 Nov;57(11):890-897. [doi: 10.1097/MLR.0000000000001185] [Medline: 31415337]
38. Kroenke K, Stump TE, Chen CX, Kean J, Bair MJ, Damush TM, et al. Minimally important differences and severity thresholds are estimated for the PROMIS depression scales from three randomized clinical trials. J Affect Disord 2020 Apr 01;266:100-108 [FREE Full text] [doi: 10.1016/j.jad.2020.01.101] [Medline: 32056864]
39. Demarzo M, Montero-Marin J, Puebla-Guedea M, Navarro-Gil M, Herrera-Mercadal P, Moreno-González S, et al. Efficacy of 8- and 4-session mindfulness-based interventions in a non-clinical population: a controlled study. Front Psychol 2017 Aug 8;8:1343 [FREE Full text] [doi: 10.3389/fpsyg.2017.01343] [Medline: 28848465]
40. Klatt MD, Buckworth J, Malarkey WB. Effects of low-dose mindfulness-based stress reduction (MBSR-ld) on working adults. Health Educ Behav 2009 Jun;36(3):601-614. [doi: 10.1177/1090198108317627] [Medline: 18469160]
41. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
42. Nicholas J, Proudfoot J, Parker G, Gillis I, Burckhardt R, Manicavasagar V, et al. The ins and outs of an online bipolar education program: a study of program attrition. J Med Internet Res 2010 Dec 19;12(5):e57 [FREE Full text] [doi: 10.2196/jmir.1450] [Medline: 21169169]
43. Al-Asadi AM, Klein B, Meyer D. Posttreatment attrition and its predictors, attrition bias, and treatment efficacy of the anxiety online programs. J Med Internet Res 2014 Oct 14;16(10):e232 [FREE Full text] [doi: 10.2196/jmir.3513] [Medline: 25316533]
44. Mermelstein LC, Garske JP. A brief mindfulness intervention for college student binge drinkers: a pilot study. Psychol Addict Behav 2015 Jun;29(2):259-269. [doi: 10.1037/adb0000040] [Medline: 25402833]
45. Be Mindful. Wellmind Health. 2022. URL: https://www.bemindfulonline.com/ [accessed 2020-09-16]
46. Querstret D, Cropley M, Fife-Schaw C. The effects of an online mindfulness intervention on perceived stress, depression and anxiety in a non-clinical sample: a randomised waitlist control trial. Mindfulness (N Y) 2018;9(6):1825-1836 [FREE Full text] [doi: 10.1007/s12671-018-0925-0] [Medline: 30524514]

## Abbreviations

**GEE:** generalized estimating equation
**MBCT:** mindfulness-based cognitive therapy

**MBSR:** mindfulness-based stress reduction
**PPRN:** patient-powered research network
**SMD:** standardized mean difference
**WHO-5:** World Health Organization—Five Well-Being Index

---

*Edited by A Mavragani; submitted 14.12.21; peer-reviewed by S O'Sullivan, P Wicks; comments to author 15.03.22; revised version received 03.05.22; accepted 13.05.22; published 12.09.22.*

*Please cite as:*
*Sylvia LG, Lunn MR, Obedin-Maliver J, McBurney RN, Nowell WB, Nosheny RL, Mularski RA, Long MD, Merkel PA, Pletcher MJ, Tovey RE, Scalchunes C, Sutphen R, Martin AS, Horn EJ, O'Boyle M, Pitch L, Seid M, Redline S, Greenebaum S, George N, French NJ, Faria CM, Puvanich N, Rabideau DJ, Selvaggi CA, Yu C, Faraone SV, Venkatachalam S, McCall D, Terry SF, Deckersbach T, Nierenberg AA*
*Web-Based Mindfulness-Based Interventions for Well-being: Randomized Comparative Effectiveness Trial*
*J Med Internet Res 2022;24(9):e35620*
*URL: https://www.jmir.org/2022/9/e35620*
*doi:10.2196/35620*
*PMID:36094813*

Original Paper

# Adapting the Eliminating Medications Through Patient Ownership of End Results Protocol to Promote Benzodiazepine Cessation Among US Military Veterans: Focus Group Study With US Military Veterans and National Veterans Health Administration Leaders

Michael A Cucciare[1,2,3], PhD; Traci H Abraham[1,2,3*], PhD; Lakiesha Kemp[1*], MA; Penny White[2*], BS; Kathy Marchant[1*], BSN; Hildi J Hagedorn[4,5*], PhD; Keith Humphreys[6,7*], PhD

[1]Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Affairs Healthcare System, North Little Rock, AR, United States

[2]Veterans Affairs South Central Mental Illness Research, Education and Clinical Center, Central Arkansas Veterans Healthcare System, North Little Rock, AR, United States

[3]Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, AR, United States

[4]Center for Care Delivery and Outcomes Research, Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, United States

[5]Department of Psychiatry, University of Minnesota Medical School, Minneapolis, MN, United States

[6]Center for Innovation to Implementation, Veterans Affairs Palo Alto Health Care System, Menlo Park, CA, United States

[7]Department of Psychiatry, Stanford University, Menlo Park, CA, United States

*these authors contributed equally

**Corresponding Author:**
Michael A Cucciare, PhD
Center for Mental Healthcare and Outcomes Research
Central Arkansas Veterans Affairs Healthcare System
2200 Fort Roots Drive
North Little Rock, AR, 72114
United States
Phone: 1 526 8179
Email: MACucciare@uams.edu

## Abstract

**Background:** Long-term dependence on prescribed benzodiazepines is a public health problem. Eliminating Medications Through Patient Ownership of End Results (EMPOWER) is a promising self-management intervention, delivered directly to patients as a printed booklet, that is effective in promoting benzodiazepine reduction and cessation in older adults. EMPOWER has high potential to benefit large health care systems such as the US Veterans Health Administration (VHA), which cares for many veterans who use benzodiazepines for extended periods.

**Objective:** We aimed to adapt the original EMPOWER booklet materials for electronic delivery and for use among US military veterans receiving VHA care who were long-term benzodiazepine users.

**Methods:** We used elements of Analysis, Design, Development, Implementation, and Evaluation, a framework commonly used in the field of instructional design, to guide a qualitative approach to iteratively adapting EMPOWER Electronic Delivery (EMPOWER-ED). We conducted 3 waves of focus groups with the same 2 groups of VHA stakeholders. Stakeholders were VHA-enrolled veterans (n=16) with medical chart evidence of long-term benzodiazepine use and national VHA leaders (n=7) with expertise in setting VHA policy for prescription benzodiazepine use and developing electronically delivered educational tools for veterans. Qualitative data collected from each wave of focus groups were analyzed using template analysis.

**Results:** Themes that emerged from the initial focus groups included veterans' anxiety about self-tapering from benzodiazepines and prior negative experiences attempting to self-taper without support. Participants also provided feedback on the protocol's look and feel, educational content, the tapering protocol, and website functionality; for example, feedback from policy leaders included listing, on the cover page, the most commonly prescribed benzodiazepines to ensure that veterans were aware of medications that qualify for self-taper using the EMPOWER-ED protocol. Both groups of stakeholders identified the importance of having access to supportive resources to help veterans manage sleep and anxiety in the absence of taking benzodiazepines. Both groups also emphasized the importance of ensuring that the self-taper could be personalized and that the taper instructions were clear. The policy leaders emphasized the importance of encouraging veterans to notify their provider of their decision to self-taper to help facilitate provider assistance, if needed, with the taper process and to help prevent medication stockpiling.

**Conclusions:** EMPOWER-ED is the first direct-to-patient electronically delivered protocol designed to help US military veterans self-taper from long-term benzodiazepine use. We used the Analysis, Design, Development, Implementation, and Evaluation framework to guide the successful adaption of the original EMPOWER booklet for use with this population and for electronic delivery. The next step in this line of research is to evaluate EMPOWER-ED in a randomized controlled trial.

***(J Med Internet Res 2022;24(9):e35514)*** doi:10.2196/35514

**KEYWORDS**

self-taper; electronically delivered self-help; long-term benzodiazepine use; US military veterans; mobile phone

# *Introduction*

## Background

Long-term dependence on prescribed benzodiazepines is a public health problem in multiple countries, including the United Kingdom and the United States [1-4]. Benzodiazepines can provide short-term (<4 weeks) benefits, but they also convey significant risks that worsen over time [5]. Adverse outcomes of long-term benzodiazepine use can include cognitive decline, falls, motor vehicle accidents, benzodiazepine dependence, and opioid-benzodiazepine overdose [6-8]. Furthermore, although benzodiazepines can be helpful for improving insomnia or severe anxiety in the short term, over the long term, they can make these conditions worse [9]. Therefore, interventions are needed to help people discontinue long-term use of these widely prescribed drugs. This paper describes an iterative, qualitative approach to adapting a direct-to-patient intervention—Eliminating Medications Through Patient Ownership of End Results (EMPOWER)—for electronic delivery [10,11]. To better promote benzodiazepine cessation specifically among US military veterans, this project also tailored EMPOWER to this population. US military veterans were the focus because they are at high risk for longer-term benzodiazepine use [12] and associated health consequences [13].

Cognitive behavioral therapy delivered by a provider, coupled with medication tapering, is effective for reducing long-term benzodiazepine use. Among adults with generalized anxiety disorder, those assigned to receive 12 weeks of cognitive behavioral therapy with medication tapering were more likely to discontinue benzodiazepine use after the intervention than control participants (75% vs 37%, respectively) [14]. Although cognitive behavioral therapy delivered by a provider can reduce benzodiazepine use [14,15], limits on patient willingness and resources, including access, time, and cost, pose challenges to delivering professionally administered in-person benzodiazepine cessation interventions to large populations.

Brief interventions delivered in the primary care setting can also lead to significant reductions in benzodiazepine use and are arguably more feasible to deliver to large populations of adults than more intensive interventions such as cognitive behavioral therapy; for example, in a randomized controlled trial, adult long-term users of benzodiazepines received either a structured primary care intervention consisting of education on the risks of long-term benzodiazepine use, a self-help leaflet to improve sleep, and a gradual medication taper with primary care provider follow-up or a written tailored dose reduction schedule or standard care [16]. At 12-month follow-up, there were no differences between the 2 structured-intervention conditions (provider follow-up or written taper schedule), with results showing that more patients in the 2 brief structured-intervention groups discontinued their benzodiazepines than those assigned to standard care (45% vs 15%, respectively) [16]. The findings highlight the feasibility and effectiveness of using direct-to-patient interventions to reduce long-term benzodiazepine use, although, as with psychotherapy, brief interventions depend on trained professionals to deliver them and patients to be available in person to receive them.

Technology can make possible a less costly and more scalable approach: providing an easily accessible, direct-to-patient intervention that educates individuals about the risks of taking benzodiazepines over the long-term encourages reduction or cessation of benzodiazepine use and provides tools to help patients taper on their own or in consultation with a provider. Mounting such an intervention on the internet brings further advantages: interventions that do not require patients to travel and that can be accessed at any time and from virtually any place could be appealing to those who might not have the desire or access to use in-person care in a clinic setting.

EMPOWER is a promising direct-to-patient self-management intervention developed by Canadian researcher Dr Cara Tannenbaum [10,11]. EMPOWER provides information about the potential risks and harms of long-term benzodiazepine use and presents alternative, effective options for reducing insomnia or anxiety and help with self-tapering. In the EMPOWER study, older adults (n=148) with long-term benzodiazepine use were mailed the EMPOWER booklet and, compared with controls (n=155), were 8 times more likely to discontinue their use of benzodiazepines [10]. Impressively, these results were obtained without direct care by a clinical professional. If EMPOWER were to be adapted for electronic delivery, such as being accessible by desktop computer, tablet computer, or mobile phone and found to be effective in this format, its reach could be further expanded.

The US Veterans Health Administration (VHA) has significant potential to adapt scalable, effective self-management interventions to help veterans reduce their long-term benzodiazepine use. The VHA is a government-financed health care system that offers comprehensive care to >9 million individuals with prior service in the US military. At the VHA, 355,298 veterans were prescribed benzodiazepines in the fiscal year 2016, almost two-thirds (63.6%) of whom took them for ≥3 months [12]. The VHA patient population includes many older adults [17] and many who take prescribed opioids [18], which increases the health risks of long-term benzodiazepine use.

## Objectives

In this paper, we describe an iterative, qualitative approach used to adapt the original EMPOWER protocol for electronic delivery (EMPOWER-ED) and for use among veterans receiving care at the VHA with long-term benzodiazepine use. To achieve these 2 objectives, we used elements of Analysis, Design, Development, Implementation, and Evaluation (ADDIE), a framework commonly used in the field of instructional design, to iteratively develop EMPOWER-ED [19]. The ADDIE framework uses an iterative approach to identify instructional needs and objectives and to obtain feedback from key stakeholders on initial drafts of instructional content and functioning website prototypes. Through formative evaluation, the ADDIE framework can also help to determine program usability, acceptability, and potential for effectiveness [19,20].

For this study, we used the analysis, design, development, and implementation elements of the ADDIE framework to guide our iterative, qualitative approach to adapting EMPOWER for use among veterans and for electronic delivery. To achieve this, we conducted 3 waves of focus group discussions with 2 groups of key VHA stakeholders. Stakeholders were VHA-enrolled veterans who were long-term users of benzodiazepines and national VHA leaders with expertise in setting VHA policy for prescription medication use, including for benzodiazepines, and in developing electronically delivered self-management tools for veterans. In this paper, we present an overview of qualitative findings from the focus group discussions, including themes representing stakeholder opinions and experiences of using benzodiazepines and attempting to taper as well as specific recommendations for adapting EMPOWER for use among veterans and for electronic delivery. We also provide example images of EMPOWER-ED to highlight key adaptations and their rationale for this population.

# Methods

## Participants and Recruitment

The principal investigators (PIs; MAC and KH) of this study chose to form 2 focus groups: one comprised US military veterans and the other comprised national VHA policy leaders. These 2 types of stakeholders were chosen because of their relevance to the aim of the study, which was to adapt an intervention protocol for the US veteran population and for electronic delivery. Stakeholders were also chosen because the study PIs (MAC and KH), who are both middle-aged White male psychologists and senior health services researchers at the VHA, had access to these populations of individuals at their respective sites [21].

We recruited veterans (n=16) with at least one primary care visit in the prior year at one of the 2 study sites (Veterans Affairs Palo Alto Health Care System in California and Central Arkansas Veterans Healthcare System in Arkansas) and with electronic medical record evidence of long-term or ≥3 months of continuous prescription benzodiazepine use. The purpose of recruiting veterans was to help ensure that EMPOWER-ED content was acceptable to this patient population and reflected their experiences and care needs. Potentially eligible veterans were identified in the corporate data warehouse, which contains data from the VHA's electronic health record. Veterans meeting initial study eligibility, such as having ≥1 primary care visit in the last year and ≥3 months of a continuous benzodiazepine prescription, were identified and randomly sent letters inviting them to participate in the study.

We also recruited national VHA leaders (n=7) with expertise in VHA care policy for prescription medication use, including benzodiazepines, and in the development and implementation of a web-delivered direct-to-patient educational intervention. The inclusion of national VHA policy leaders helped to ensure that the content of EMPOWER-ED was consistent with VHA priorities and that information was conveyed using up-to-date instructional design techniques. As there is a limited number of national VHA policy leaders with expertise in the aforementioned areas of interest, we used nonrandom selection, which involved identification by the 2 study PIs (MAC and KH) to invite these individuals to participate. We also attempted to recruit VHA primary care providers at the 2 study sites but were unsuccessful.

Each set of stakeholders was invited to participate in all 3 waves of focus group discussions over a 12-month period lasting on average 60 minutes; 16, 14, and 9 veterans participated in waves 1, 2, and 3, respectively, whereas 6, 6, and 5 national VHA leaders participated in waves 1, 2, and 3, respectively. Each stakeholder type participated in a focus group discussion with participants of the same type. Focus groups with veterans took place between October 2020 and September 2021, whereas focus groups with national VHA policy leaders took place between March 2021 and September 2021.

## Description of the Original EMPOWER (Booklet) Protocol

The original EMPOWER protocol consists of an 8-page booklet that was mailed directly to older adults with long-term benzodiazepine use [10,11]. The booklet includes a self-assessment of the potential risks of long-term benzodiazepine use; evidence of benzodiazepine-related harms; knowledge statements designed to evoke cognitive dissonance about the safety of using benzodiazepines; education about possible drug interactions; a vignette depicting a peer who has successfully stopped using benzodiazepines to support self-efficacy to change medication use; information about equally or more effective therapeutic alternatives for managing sleep difficulties or anxiety; and recommendations and guidance for self-tapering their medication, including a taper schedule. The taper schedule consists of a 21-week protocol with daily guidance, regardless of original dose, for reducing medication use. The taper schedule also includes the recommendation for the patient to discuss their self-taper with their physician or pharmacist.

## Data Collection

Focus groups, conducted through the web, were the primary method of data collection. All focus group discussions were audio recorded to facilitate qualitative data collection and analysis. Veterans and national VHA leaders participated in focus group discussions by teleconference using Microsoft Teams, a web-based meeting portal that allows participants to view the same content simultaneously. All focus group discussions were comoderated by the 2 study PIs (MAC and KH), both of whom are clinically trained and have extensive experience in qualitative interviewing and facilitating focus group discussions. An interview guide was developed for the first wave of focus groups to ensure that the areas discussed remained roughly consistent across stakeholder groups and that all relevant topics were addressed. Constructs from the Consolidated Framework for Implementation Research (CFIR [22]) were used to inform the development of the initial interview guide (Textbox 1). The CFIR was chosen because it provides a framework of constructs that can help to identify strategies for optimizing the adoption of a new innovation. In this study, we used the CFIR constructs of intervention characteristics, which includes perceived effectiveness and relative advantage; outer setting, which includes patient care needs and available resources; and characteristics of individuals, which includes knowledge, attitudes, and perspectives toward change [22], to guide the wave 1 focus group discussions. The interview questions for waves 2 and 3 were shorter, less formal, and focused on obtaining feedback on each subsequent draft of the EMPOWER-ED website. Example questions for waves 2 and 3 included “To what extent were your suggestions implemented in the current draft as you envisioned?” “To what extent are draft materials clear and functioning properly?” In addition, the study’s lead qualitative researcher (THA) recorded observations during each wave of focus group discussions. The observations focused on group dynamics and instances in which participants diverged from the questions asked by the comoderators (MAC and KH), including the content and duration of these interactions.

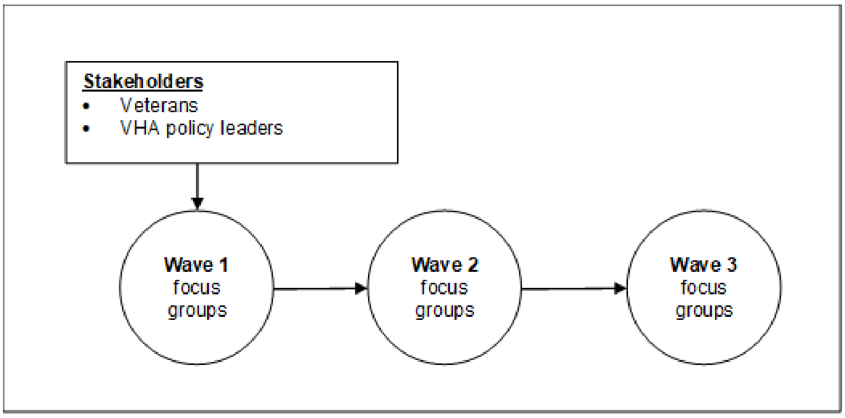
The same stakeholders were invited to participate in all 3 waves of focus group discussions (Figure 1). This facilitated consistency of feedback over time because participants were aware of the recommendations provided during prior waves and could provide feedback on whether recommendations were incorporated as desired into the current draft. Furthermore, for focus group waves 2 and 3, stakeholders were asked to interact with the draft website in the week before the focus group to facilitate discussions on user experience. At the start of the second and third waves of focus groups, the comoderators (MAC and KH) reviewed the intervention materials, described the adaptations made to content, and demonstrated website functionality.

**Textbox 1.** Interview guide for wave 1 focus groups. After each series of questions, assessing each Consolidated Framework for Implementation Research domain, the focus group facilitators obtained initial recommendations and suggestions for adding, deleting, and modifying elements to the Eliminating Medications Through Patient Ownership of End Results (EMPOWER) materials for adaptation for electronic delivery and to optimally assist veterans in reducing or discontinuing their long-term benzodiazepine use.

**Wave 1 focus group interview guide**

- Characteristics of EMPOWER
- “What are your thoughts about the content of EMPOWER?”
- “What are your thoughts and opinions about the educational materials? The vignette? The taper schedule?”
- “To what extent would each of these components of EMPOWER be helpful (or not) to veterans in reducing their use of benzodiazepines?”
- “What are your thoughts and opinions about the length of the EMPOWER materials?”
- “What changes (if any) would need to be made to EMPOWER to improve its usefulness for veterans?”
- Outer setting
- “To what extent do you feel that veterans using benzodiazepines would respond (or not) to this intervention?”
- “What would increase their positive response to these materials?”
- Characteristics of individuals
- “What challenges do you see veterans taking benzodiazepines might have self-tapering their medication?”

**Figure 1.** Participant flow through the 3 waves of focus groups. VHA: Veterans Health Administration.



### Ethics Approval

The study was approved by the VA Central Arkansas Veterans Health Care System Research and Development Committee and Institutional Review Board (protocol #1527634).

### Data Analysis

Rapid analytic techniques informed by Sobo et al [23] and Hamilton and Finlay [24] were used to quickly produce recommendations for adapting EMPOWER for use among veterans and for electronic delivery. The lead qualitative researcher (THA) developed a prototype summary template in a Microsoft Word document with domains related to the goals of the study; for example, the researcher captured stakeholder feedback and recommendations on the protocol’s look and feel, educational content, tapering protocol, and website functionality.

The lead qualitative researcher (THA) also captured emergent themes and created an *other* domain to record observations made during the focus group discussions, including the content and duration of instances in which participants spoke on issues among themselves.

For each wave of focus groups, the lead qualitative researcher (THA) first listened to the audio recording of the discussion and systematically populated the template categories with data. Content analysis was used to ensure that the full range of experiences, perspectives, and feedback was included in the templates [25]. Template data largely consisted of paraphrased content from the discussions, reflecting stakeholder recommendations, reactions, concerns, and questions. Particularly impactful statements were transcribed verbatim into the templates. After this initial step, the lead qualitative researcher (THA) synthesized individual templates from each wave of focus groups by stakeholder type into 1 template. In some waves, multiple focus groups took place for a specific stakeholder type (eg, 2 separate focus group discussions for different veterans) to accommodate their schedules. When this occurred, data were synthesized into 1 template for that stakeholder group for that specific wave of focus groups.

In the final step, the lead qualitative researcher (THA) synthesized the 2 templates by stakeholder group into 1 summary template containing all feedback for that wave. The researcher also carefully reviewed the *other* domain to identify any themes occurring across focus group discussions. Themes identified were mapped to the CFIR domains and included in the summary template. The comoderators (MAC and KH) of the focus groups reviewed the summary template for completeness and accuracy.

### Finalizing Adaptations

After each of the 3 waves of focus groups, the 2 project PIs (MAC and KH) met with the software development team to review the summary template, which included feedback from both sets of stakeholders. Recommendations deemed to be of high priority, such as being consistent with the original EMPOWER materials or suggested by multiple stakeholders, and feasible, such as being within the cost and time parameters needed to implement the recommendation, were identified and implemented. In waves 2 and 3, participants were asked to review the website modifications to determine whether they were implemented as suggested and provide additional feedback or recommendations.

## Results

### Overview

In the following sections, we first present the themes that emerged from the qualitative analysis, which included veterans’ anxiety about self-tapering and their experiences of attempting to self-taper without support (Textbox 2). We then present focus group participants’ feedback and recommendations on modifying the look and feel, educational content, tapering protocol, and functionality of the EMPOWER-ED website to optimize its adaptation for electronic delivery and use among veterans. Although we attempted to recruit primary care physicians to participate in the focus groups, we were unsuccessful. The primary reason reported by physicians for not participating was that primary care clinics were focusing efforts on the ongoing COVID-19 pandemic, and they had limited time to contribute to non–patient-care activities, including research.

**Textbox 2.** Emergent themes as well as feedback and recommendations on Eliminating Medications Through Patient Ownership of End Results (EMPOWER) and Eliminating Medications Through Patient Ownership of End Results Electronic Delivery (EMPOWER-ED).

**Themes that emerged from focus group participants' feedback and recommendations**

- Emergent themes
  - Veterans' anxiety about self-tapering their benzodiazepines
  - Veterans' prior negative experiences attempting to taper without effective, alternative help options for managing sleep and anxiety
- Focus group feedback on the EMPOWER and EMPOWER-ED content
  - Look and feel
    - Include images that reflect veterans' diversity
    - Develop multiple peer vignettes that include a diverse group of veterans
  - Educational content
    - Clarify names (brand and generic) of commonly prescribed benzodiazepines
    - Emphasize potential risks of longer-term benzodiazepine use
    - Include alternative help options, including websites and mobile apps, for managing sleep and anxiety
  - Tapering protocol and website functionality
    - Encourage veterans to inform their prescribing provider of their decision to taper their benzodiazepines
    - Include clear language instruction for veterans on how much medication to take each day during the taper process
    - Allow veterans to save their taper schedule on the website
    - Ensure that the taper schedule is readable on all devices, including smartphones, tablet computers, and desktop computers

## Emergent Themes

Two themes emerged from analysis of the veterans' interactions during the 3 waves of focus group discussions related to the CFIR domain Characteristics of Individuals. The first theme reflected veterans' anxiety about self-tapering, specifically concerns about how to manage sleep or anxiety without their medication. Illustrating this anxiety, a veteran stated, "It's hard to get some of those war-time experiences out of your head...so being without the medication can be scary" (focus group 3, veteran 03).

An additional theme reflected the veterans' prior negative experiences of attempting to self-taper without effective, alternative help options for managing sleep or mental health symptoms; for example, a veteran described attempting to self-taper in this context as feeling like "...falling into a bottomless pit" (focus group 2, veteran 04). Similarly, the veterans repeatedly expressed a strong perceived need for having effective resources and support during the self-taper process:

> *I believe that if anyone decides to taper off their medication, they need to have...assistance. When it is your first time being tapered off, you know you want to be counseled.* [Focus group 2, veteran 01]

## Focus Group Feedback on the EMPOWER and EMPOWER-ED Content

### *Look and Feel*

Initial focus group feedback centered on modifying the look and feel of the website, including increasing the font size when possible, using less white space, adding more color, and ensuring that the text was at a sixth grade reading level. Participants suggested that we reduce the number of benzodiazepines listed on the website's home page to those most commonly prescribed at the VHA and use generic and brand names to facilitate identification of these medications by veterans (Figures 2 and 3; modifications to the cover page included simplifying the list of *qualifying* medications that veterans may consider tapering [covering approximately 99% of the benzodiazepine prescriptions at the VHA], including generic and brand names of benzodiazepines to facilitate understanding of which medications are *eligible* for tapering and providing education about the use and risks of these medications over the longer term).

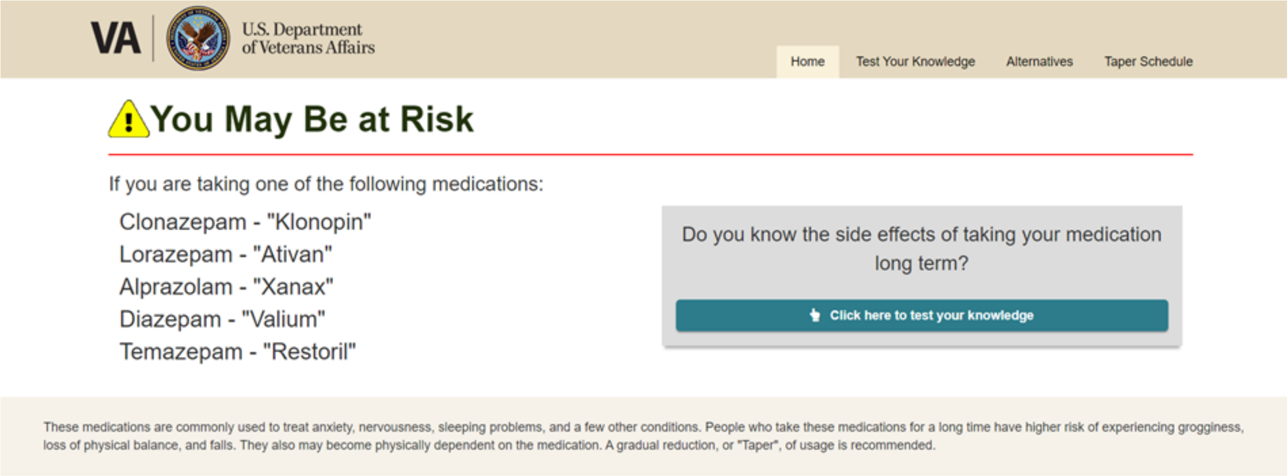
**Figure 2.** Cover page of the original Eliminating Medications Through Patient Ownership of End Results booklet.

# You May Be at Risk

You are taking one of the following sedative-hypnotic medications:

○ Alprazolam (Xanax®)
○ Chlorazepate
○ Chlordiazepoxide-amitriptyline
○ Clidinium-Chlordiazepoxide
○ Clobazam
○ Clonazepam (Rivotril®, Klonopin®)
○ Diazepam (Valium®)
○ Estazolam
○ Flurazepam
○ Loprazolam
○ Lorazepam (Ativan®)
○ Lormetazepam
○ Nitrazepam
○ Oxazepam (Serax®)
○ Quazepam
○ Temazepam (Restoril®)
○ Triazolam (Halcion®)
○ Eszopiclone (Lunesta®)
○ Zaleplon (Sonata®)
○ Zolpidem (Ambien®, Intermezzo®, Edluar®, Sublinox®, Zolpimist®)
○ Zopiclone (Imovane®, Rhovane®)

**Figure 3.** Cover page of the Eliminating Medications Through Patient Ownership of End Results Electronic Delivery website.



## You May Be at Risk

If you are taking one of the following medications:

Clonazepam - "Klonopin"
Lorazepam - "Ativan"
Alprazolam - "Xanax"
Diazepam - "Valium"
Temazepam - "Restoril"

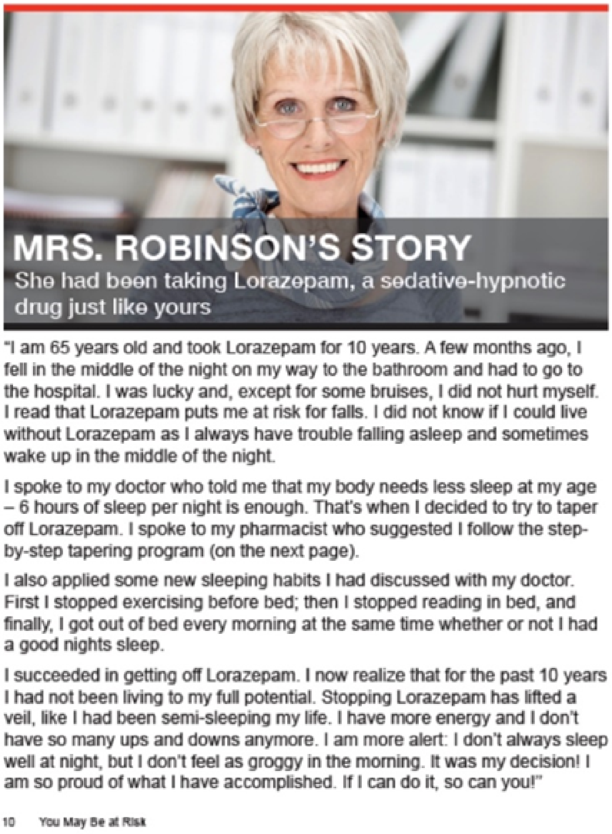
Do you know the side effects of taking your medication long term?

These medications are commonly used to treat anxiety, nervousness, sleeping problems, and a few other conditions. People who take these medications for a long time have higher risk of experiencing grogginess, loss of physical balance, and falls. They also may become physically dependent on the medication. A gradual reduction, or "Taper", of usage is recommended.

Participants recommended adding new images of veterans throughout the program to better reflect their diversity in age, gender, and race; for example, a veteran stated that "...veterans want pictures and stories that are a mix of both ages and races" (focus group 1, veteran 02). Furthermore, a VHA policy leader suggested including "...more realistic photos that don't look staged and are images of people in a more natural environment" (focus group 2, VHA policy leader 1). This suggestion was most notable for the original peer vignette that depicted an older (aged >65 years) White woman's story about discontinuing her benzodiazepine use that participants did not generally identify with. Participants recommended that we redesign the single peer vignette content to include 3 diverse peer vignettes to give veterans more choice in selecting a vignette to read or listen to, thus increasing the likelihood that veterans will identify with the content. Consistent with this recommendation, a veteran stated, "...replace reference to Mrs Robinson [woman depicted in the vignette] to different people from different backgrounds to make the materials [vignettes] more relatable to veterans" (focus group 1, veteran 03). Feedback on the peer vignettes also included reducing the length of each peer's *story*, creating an option to listen to the vignettes through audio as opposed to requiring that they be read and to ensure that veterans from diverse age groups were represented in the vignettes (Figures 4 and 5; modifications to the peer vignettes included developing 3 vignettes [replacing the original single vignette] to provide a more diverse [age and race] representation of veterans. We also provide the option for veterans to listen to each vignette in case of reading or visual impairment [not shown in image]).

In the third wave of focus groups, participants *validated* the final version of the EMPOWER-ED website by indicating that the final images, text, and peer vignettes resonated with their experiences and care needs and reflected the racial, ethnic, and gender diversity of US military veterans.

**Figure 4.** Vignette depicting peers who have successfully discontinued benzodiazepines—the original Eliminating Medications Through Patient Ownership of End Results booklet.



MRS. ROBINSON'S STORY

She had been taking Lorazepam, a sedative-hypnotic drug just like yours

"I am 65 years old and took Lorazepam for 10 years. A few months ago, I fell in the middle of the night on my way to the bathroom and had to go to the hospital. I was lucky and, except for some bruises, I did not hurt myself. I read that Lorazepam puts me at risk for falls. I did not know if I could live without Lorazepam as I always have trouble falling asleep and sometimes wake up in the middle of the night.

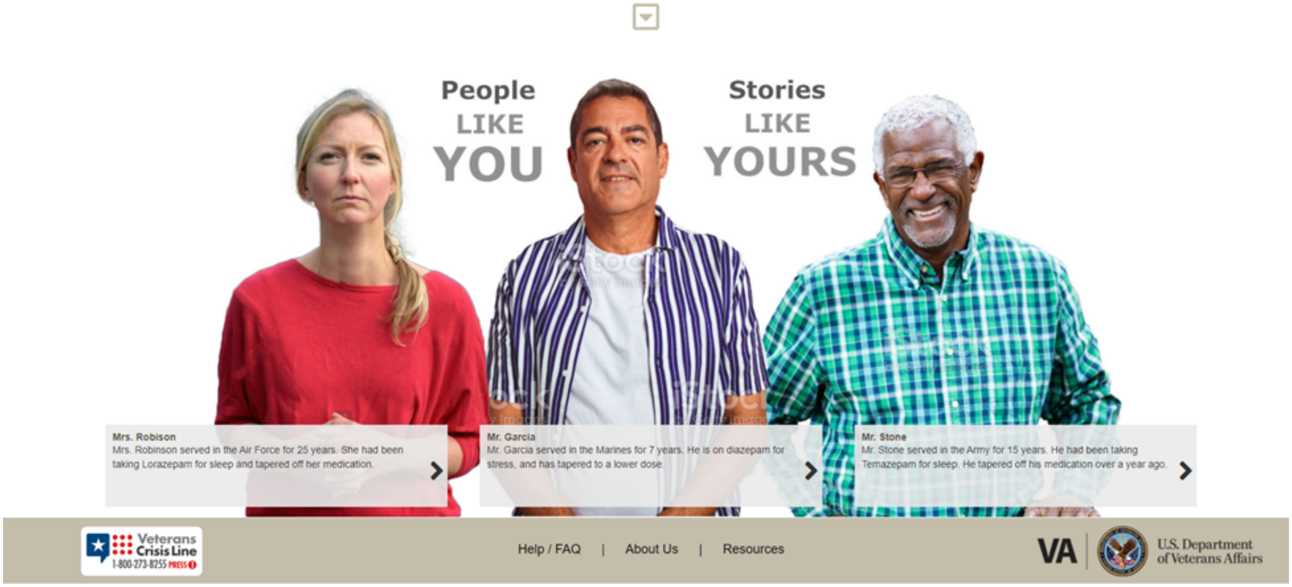
I spoke to my doctor who told me that my body needs less sleep at my age – 6 hours of sleep per night is enough. That's when I decided to try to taper off Lorazepam. I spoke to my pharmacist who suggested I follow the step-by-step tapering program (on the next page).

I also applied some new sleeping habits I had discussed with my doctor. First I stopped exercising before bed; then I stopped reading in bed, and finally, I got out of bed every morning at the same time whether or not I had a good nights sleep.

I succeeded in getting off Lorazepam. I now realize that for the past 10 years I had not been living to my full potential. Stopping Lorazepam has lifted a veil, like I had been semi-sleeping my life. I have more energy and I don't have so many ups and downs anymore. I am more alert: I don't always sleep well at night, but I don't feel as groggy in the morning. It was my decision! I am so proud of what I have accomplished. If I can do it, so can you!"

10 You May Be at Risk

**Figure 5.** Vignette depicting peers who have successfully discontinued benzodiazepines—the Eliminating Medications Through Patient Ownership of End Results Electronic Delivery website.



**Mrs. Robison**
Mrs. Robinson served in the Air Force for 25 years. She had been taking Lorazepam for sleep and tapered off her medication.

**Mr. Garcia**
Mr. Garcia served in the Marines for 7 years. He is on diazepam for stress, and has tapered to a lower dose.

**Mr. Stone**
Mr. Stone served in the Army for 15 years. He had been taking Temazepam for sleep. He tapered off his medication over a year ago.

### *Educational Content*

Recommendations regarding the educational content included providing a list of benzodiazepines that are included in the VHA's National Formulary, using generic terms for medications to educate veterans on which medications could *qualify* for the self-taper, and emphasizing the potential risks associated with longer-term benzodiazepine use and describing those risks, including memory and concentration problems, daytime fatigue, falls, fractures, and motor vehicle accidents, in understandable language. Indeed, a VHA policy leader recommended that we "...organize the list of benzos by most common [commonly prescribed at the VHA] on the first page using large font" (focus group 2, VHA policy leader 3). Regarding how to structure content about the risks of long-term benzodiazepine use, a veteran stated, "...include references to risks that are more common to veterans like drinking in combo with the meds and memory changes and falls" (focus group 2, veteran 05). In addition, participants suggested clarifying how physical dependence on benzodiazepines can occur and to provide content that educates veterans on how an effective taper process, such as a longer taper compared with a shorter one, can minimize withdrawal symptoms.

For alternative help options to medication, participants recommended that we embed links to VHA-developed websites that consist of evidence-based content to improve sleep (Path to Better Sleep) and reduce anxiety (Moving Forward: Overcoming Life's Challenges). One veteran stated as follows:

> *...you need additional support during the tapering process. You can get anything you want on the street, so just taking them off is not going to work. I think that just a little help for them [veterans]...would be good. Maybe there could be a support group that is geared toward benzodiazepine cessation.* [Focus group 1, veteran 03]

In addition, participants highly recommended adding "...a link to a website or mobile app that talks about yoga, some kind of meditation, things of that nature" (focus group 2, VHA policy leader 4). As a result, we added VHA help options that can be accessed by mobile phone, such as CBT-i Coach, Insomnia Coach, PTSD Coach, Mindfulness Coach, Beyond MST, VetChange (coping with posttraumatic stress disorder), and Virtual Hope Box (relaxation and stress-coping tools; Figures 6 and 7; modifications to the alternative help and care options section included adding stakeholder [veteran and national VHA policy leader]–suggested website and mobile apps developed by the VHA to help veterans better manage and cope with common mental health comorbidities, eg, posttraumatic stress disorder and anxiety, and other experiences that can cause stress and anxiety as well as learn coping skills for managing stress as well as problem-solving and parenting skills). On the basis of participant feedback, we also clarified which alternative help options are downloadable when using a smartphone, tablet computer, and desktop computer. Furthermore, it was suggested that we add the VHA Crisis Line and links to documents that provide psychoeducation on basic sleep hygiene. In the third wave of focus groups, the veterans *validated* these changes and reported that the modifications to the educational content were clear and enhanced their readability and that the alternative help options included reflected their care needs.

**Figure 6.** Alternative care and help options for managing anxiety without medications—the original Eliminating Medications Through Patient Ownership of End Results booklet.

ALTERNATIVES

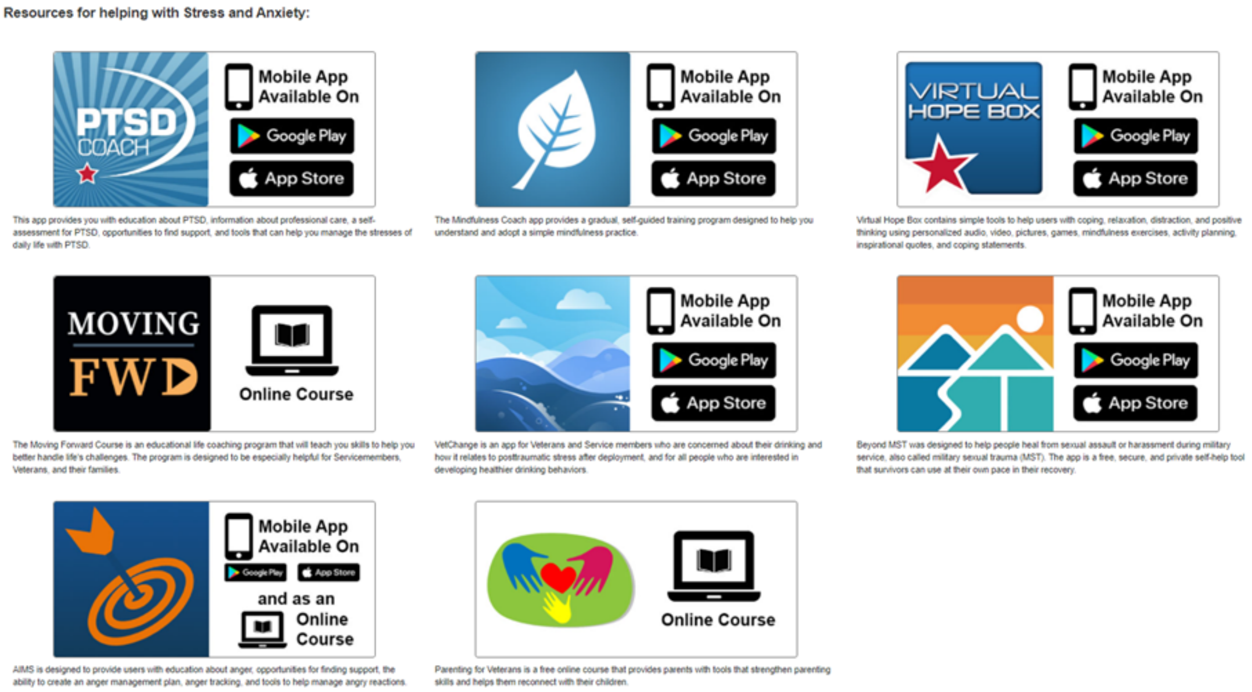
If you are taking this sedative-hypnotic drug to help reduce your anxiety:

**There are other solutions to deal with your stress and anxiety.**

- Talking to a therapist is a good way to help you work out stressful situations and talk about what makes you anxious.
- Support groups help to relieve your stress and make you feel you are not alone.
- Try relaxation techniques like stretching, yoga, massage, meditation or tai chi that can help relieve you of everyday stress and help you work through your anxiety.
- Talk to your doctor about other anti-anxiety medications that have less serious side effects.

You May Be at Risk 9

**Figure 7.** Alternative care and help options for managing anxiety without medications—the Eliminating Medications Through Patient Ownership of End Results Electronic Delivery website.



Resources for helping with Stress and Anxiety:

This app provides you with education about PTSD, information about professional care, a self-assessment for PTSD, opportunities to find support, and tools that can help you manage the stresses of daily life with PTSD.

The Mindfulness Coach app provides a gradual, self-guided training program designed to help you understand and adopt a simple mindfulness practice.

Virtual Hope Box contains simple tools to help users with coping, relaxation, distraction, and positive thinking using personalized audio, video, pictures, games, mindfulness exercises, activity planning, inspirational quotes, and coping statements.

The Moving Forward Course is an educational life coaching program that will teach you skills to help you better handle life's challenges. The program is designed to be especially helpful for Servicemembers, Veterans, and their families.

VetChange is an app for Veterans and Service members who are concerned about their drinking and how it relates to posttraumatic stress after deployment, and for all people who are interested in developing healthier drinking behaviors.

Beyond MST was designed to help people heal from sexual assault or harassment during military service, also called military sexual trauma (MST). The app is a free, secure, and private self-help tool that survivors can use at their own pace in their recovery.

AIMS is designed to provide users with education about anger, opportunities for finding support, the ability to create an anger management plan, anger tracking, and tools to help manage angry reactions.

Parenting for Veterans is a free online course that provides parents with tools that strengthen parenting skills and helps them reconnect with their children.

### *Tapering Protocol and Website Functionality*

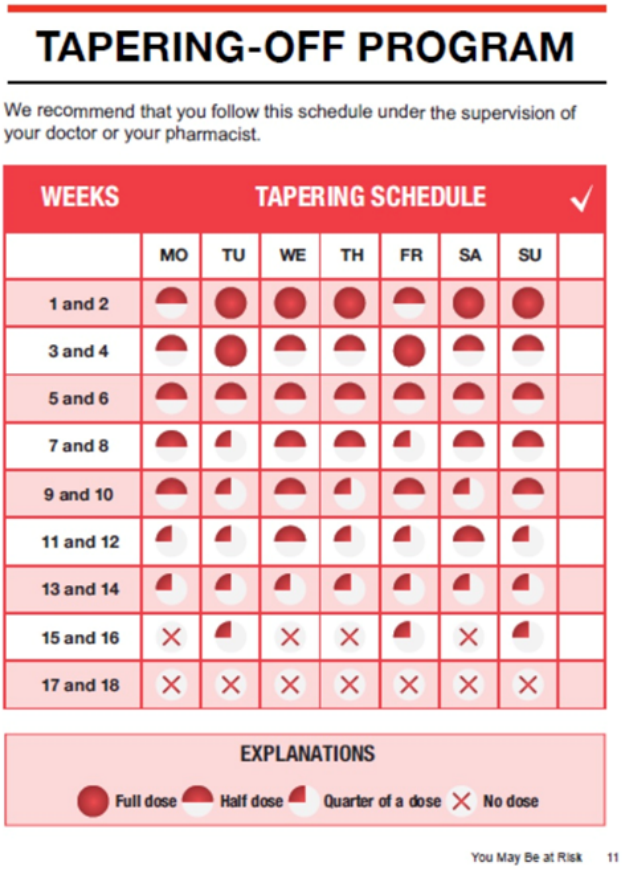
Participants suggested that we emphasize that veterans who choose to self-taper should inform their provider of their decision. A veteran summarized the potential challenge of tapering off benzodiazepines and the importance of a careful, slow taper: "I think it would be a very slow process and then you have to be very careful...you will have to set up a [taper] system where if you are going to get off of them, you need to write a note that reminds you what you need to take [each day]" (focus group 1, veteran 03). Furthermore, participants recommended that we emphasize that veterans consider speaking with their prescribing provider or a pharmacist if they need assistance while tapering, provide a link to their local VHA pharmacy to obtain a pill cutter, and more prominently display their daily taper dose provided on the taper schedule.

Participants also recommended using simpler language such as *reduce* versus *taper* and clarifying that the daily recommended medication dose on the taper schedule is what you take each time you take your medication as opposed to representing the total dosage for that day (Figures 8 and 9; modifications to the taper schedule included making the key depicting the amount of medication to take each day more prominent. We also provide the option for veterans to personalize the taper schedule to accommodate their daily dose and provide written instruction for how much medication to take each day. The program also allows veterans to verify that their daily dose is correct before it generates the personalized taper schedule [not shown in image]). Highlighting this recommendation, a VHA policy leader stated, "...be more explicit about what the person should take, including how many times a day are you taking this dosage and then integrate that into the tapering program" (focus group 2, VHA policy leader 1). Other suggestions included allowing veterans to customize their taper schedule by entering their taper start date and current dosage, including how much they should take each time they take their medication as well as frequency of daily pill taking, implementing an *accuracy check* to ensure that veterans have correctly entered their daily dosage, and allowing veterans to be able to save their taper schedule in the event it needs to be reprinted.

Feedback on website functionality consisted of fixing observed *bugs*, including enhancing the readability of the taper schedule when viewing on a tablet computer or smartphone and allowing users to move freely throughout the website as opposed to requiring that veterans complete a section before advancing to the next.

During the third wave of focus groups, participants *validated* changes made to the taper schedule and reported that the modifications made increased its accuracy, understandability, and potential effectiveness to help veterans taper their benzodiazepines. In addition, participants reported that the modifications made to the website's functionality reflected their experiences and care needs.

**Figure 8.** Taper schedule included in the original Eliminating Medications Through Patient Ownership of End Results booklet.



TAPERING-OFF PROGRAM

We recommend that you follow this schedule under the supervision of your doctor or your pharmacist.

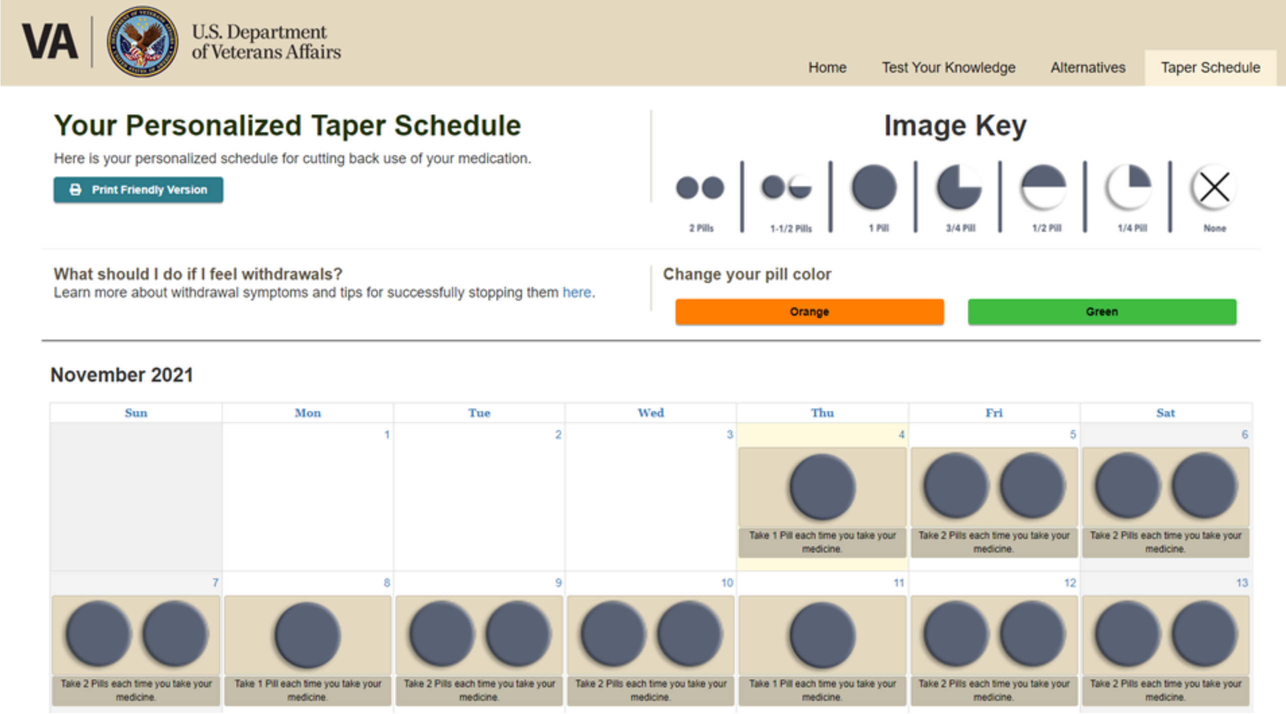
| WEEKS | TAPERING SCHEDULE | | | | | | | ✓ |
|---|---|---|---|---|---|---|---|---|
| | MO | TU | WE | TH | FR | SA | SU | |
| 1 and 2 | ◓ | ● | ● | ● | ◓ | ● | ● | |
| 3 and 4 | ◓ | ● | ◓ | ◓ | ● | ◓ | ◓ | |
| 5 and 6 | ◓ | ◓ | ◓ | ◓ | ◓ | ◓ | ◓ | |
| 7 and 8 | ◓ | ◔ | ◓ | ◓ | ◔ | ◓ | ◓ | |
| 9 and 10 | ◓ | ◔ | ◓ | ◔ | ◓ | ◔ | ◓ | |
| 11 and 12 | ◔ | ◔ | ◓ | ◔ | ◔ | ◓ | ◔ | |
| 13 and 14 | ◔ | ◔ | ◔ | ◔ | ◔ | ◔ | ◔ | |
| 15 and 16 | ✕ | ◔ | ✕ | ✕ | ◔ | ✕ | ◔ | |
| 17 and 18 | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ | ✕ | |

EXPLANATIONS

● Full dose ◓ Half dose ◔ Quarter of a dose ✕ No dose

You May Be at Risk 11

**Figure 9.** Taper schedule included in the Eliminating Medications Through Patient Ownership of End Results Electronic Delivery website.



Your Personalized Taper Schedule

Here is your personalized schedule for cutting back use of your medication.

Image Key

What should I do if I feel withdrawals?

Learn more about withdrawal symptoms and tips for successfully stopping them here.

Change your pill color

November 2021

| Sun | Mon | Tue | Wed | Thu | Fri | Sat |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4<br>Take 1 Pill each time you take your medicine. | 5<br>Take 2 Pills each time you take your medicine. | 6<br>Take 2 Pills each time you take your medicine. |
| 7<br>Take 2 Pills each time you take your medicine. | 8<br>Take 1 Pill each time you take your medicine. | 9<br>Take 2 Pills each time you take your medicine. | 10<br>Take 2 Pills each time you take your medicine. | 11<br>Take 1 Pill each time you take your medicine. | 12<br>Take 2 Pills each time you take your medicine. | 13<br>Take 2 Pills each time you take your medicine. |

## Discussion

Findings from the 3 waves of focus groups with 2 distinct types of stakeholders revealed emergent themes, including veterans' concerns about self-tapering their benzodiazepines and prior negative experiences of attempting to self-taper these medications without support. We also received feedback and recommendations on modifying the EMPOWER protocol, including on its look and feel, educational content, tapering protocol, and website functionality, which is described in more detail in the following section.

### Principal Findings and Comparison With Prior Work

Engaging 2 groups of stakeholders in each of 3 waves of focus groups helped us to identify veterans' concerns about medication tapering and needed adaptations and modifications to the EMPOWER materials to improve their usability and potential effectiveness. Taking into account the needs and preferences of stakeholders or *end users* is critical to designing engaging and effective patient self-management tools [26]; for example, participants from both stakeholder groups emphasized the importance of having access to supportive resources and help options to help manage sleep difficulties or anxiety in the absence of taking benzodiazepines. This resulted in adding several VHA-developed mobile apps and websites with evidence-based content for coping with sleep difficulties and mental health comorbidities, including anxiety and stress. Although EMPOWER-ED is designed to support the self-taper of benzodiazepines, feedback from policy leaders provided valuable insight into the importance of encouraging veterans to discuss their decision to reduce or stop using their medication with their provider. This would facilitate providers being informed of the veterans' intent and providing assistance when needed, as well as reducing the potential for *stockpiling* medication by reducing automatic medication prescriptions being sent to the veterans' homes.

As focus group participants were able to interact with the EMPOWER-ED website between focus groups, they also identified potential usability challenges. Feedback from policy leaders during this process directly informed the need to list, on the cover page, the most commonly prescribed benzodiazepines at the VHA so that veterans would be aware of medications that *qualify* for tapering. Both groups of stakeholders emphasized the need to ensure that the taper schedule was clear, including ensuring that veterans' daily dosage was accurate, before generating their personalized taper schedule. Furthermore, both stakeholder groups suggested strongly that veterans should be able to access their taper schedule in a readable form on a smartphone and tablet computer. This latter feature was important because our intention was to design EMPOWER-ED to be accessible through any electronic device. As a result of these modifications, both groups of stakeholders approved the final version of the EMPOWER-ED website, reporting that the content reflected veterans' care needs; was accurate, understandable, and easy to navigate; and functioned well. Indeed, providing high-quality educational content and having easy and well-functioning navigation can help facilitate high engagement with patient self-management websites [27].

By including both stakeholder groups, we were able to leverage different perspectives and opinions about EMPOWER-ED to optimize its potential uptake and effectiveness among this population of veterans. We found it particularly beneficial to attend to instances in which veterans went *off topic* to share their experiences of attempting to self-taper from benzodiazepines without support. Allowing veterans to share their prior experiences of attempting—and failing—to self-taper from benzodiazepines underscored the vital importance of additional support during the tapering process, resulting in an end product that more accurately reflected veterans' preferences and needs. Thus, although recommendations from VHA policy leaders were critical for identifying educational content and optimizing their clarity, the veteran focus groups resulted in adaptations that could ultimately enhance uptake of the intervention among veterans interested in making a change in their long-term use of benzodiazepines. Having both stakeholder groups provide input helped the team to develop a version of EMPOWER-ED that was acceptable and personalized to the needs of the stakeholder groups—all of which are considered key features of self-management websites that promote a positive user experience and clinical outcomes [28].

### Strengths and Limitations

A strength of this study is that it used an instructional design framework (ADDIE) to guide an iterative, qualitative approach to adapting EMPOWER [10], a protocol for promoting benzodiazepine cessation, for veterans with long-term benzodiazepine use and for electronic delivery (EMPOWER-ED). The ADDIE framework embraces an iterative formative evaluation [29] focused on assessing individuals' care needs, the need for adaptations and modifications, and the usability and usefulness of new e-learning products. The ADDIE framework also emphasizes the importance of obtaining key stakeholder feedback to inform the design and development of educational protocols. Advantages of including stakeholder feedback are having an end product that accurately represents the needs of stakeholders, is acceptable and feasible for use, and is easy to use and navigate [30]—all characteristics that can have a substantial effect on the uptake and effectiveness of e-learning tools [31]. Consequently, the ADDIE framework is increasingly being used in health care to design e-learning products that aim to change end-user behavior; for example, the ADDIE framework has been used to guide the development of e-learning protocols that help people to connect to supportive employment [30], adjust to injury and cope with pain [32], and improve mobility and physical functioning [31].

This study also includes some limitations. First, we did not conduct formal usability testing on the EMPOWER-ED website. As a result, we may have overlooked important issues concerning its usability and functioning with respect to our target population. However, before focus group waves 2 and 3, stakeholders were emailed a link to the most recent iteration of the draft website with the request that they interact with it to facilitate discussion in the subsequent focus group. During the focus groups, it was clear during discussions that some of the participants had interacted with the website, but this was not systematically assessed. Second, despite our attempts, we were not successful in recruiting VHA primary care physicians to give us feedback on the development of EMPOWER-ED. Thus, we may have failed to identify important design or functional modifications that may be important to physicians when helping veterans to self-taper from these medications.

### Conclusions and Next Steps

To our knowledge, EMPOWER-ED is the first direct-to-patient electronically delivered protocol designed to help US military veterans with evidence of long-term benzodiazepine use self-taper their benzodiazepines. In this first step of our research program, our team followed the ADDIE framework to guide the successful adaption of the original EMPOWER booklet [10] for use with this population and for electronic delivery. This resulted in the successful development of EMPOWER-ED. The next step in this line of research is to formally evaluate the effectiveness of EMPOWER-ED in a randomized controlled trial, including examining its impact on veterans' benzodiazepine use, anxiety symptoms, sleep quality, and overall health.

## Acknowledgments

The authors wish to thank Brian Nelson and Summer Jones for their technical expertise in developing Eliminating Medications Through Patient Ownership of End Results Electronic Delivery. The authors would also like to thank the US Department of Veterans Affairs Health Services Research and Development Service for supporting this work (IIR 18-026; to KH and MAC). The views expressed in this paper are those of the authors and do not necessarily reflect the official position of the US Department of Veterans Affairs.

## Conflicts of Interest

None declared.

## References

1. Steinman MA, Low M, Balicer RD, Shadmi E. Epidemic use of benzodiazepines among older adults in Israel: epidemiology and leverage points for improvement. J Gen Intern Med 2017 Aug 3;32(8):891-899 [FREE Full text] [doi: 10.1007/s11606-017-4059-1] [Medline: 28470549]
2. Taylor S, Annand F, Burkinshaw P, Greaves F, Kelleher M, Knight J, et al. Dependence and withdrawal associated with some prescribed medicines: an evidence review. Public Health England. 2019. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/940255/PHE_PMR_report_Dec2020.pdf [accessed 2022-08-30]
3. Bachhuber MA, Hennessy S, Cunningham CO, Starrels JL. Increasing benzodiazepine prescriptions and overdose mortality in the United States, 1996–2013. Am J Public Health 2016 Apr;106(4):686-688. [doi: 10.2105/ajph.2016.303061]
4. Lembke A, Papac J, Humphreys K. Our other prescription drug problem. N Engl J Med 2018 Feb 22;378(8):693-695. [doi: 10.1056/nejmp1715050]
5. Kennedy KM, O'Riordan J. Prescribing benzodiazepines in general practice. Br J Gen Pract 2019 Feb 28;69(680):152-153. [doi: 10.3399/bjgp19x701753]
6. Park TW, Larochelle MR, Saitz R, Wang N, Bernson D, Walley AY. Associations between prescribed benzodiazepines, overdose death and buprenorphine discontinuation among people receiving buprenorphine. Addiction 2020 May;115(5):924-932 [FREE Full text] [doi: 10.1111/add.14886] [Medline: 31916306]
7. Sun E, Dixit A, Humphreys K, Darnall BD, Baker L, Mackey S. Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. BMJ 2017 Mar 14;356:j760 [FREE Full text] [doi: 10.1136/bmj.j760] [Medline: 28292769]
8. Benzodiazepine risks are you aware of the possible risks from taking benzodiazepines? U.S. Department of Veterans Affairs. 2016 Oct. URL: https://www.pbm.va.gov/PBM/AcademicDetailingService/Documents/Benzodiazepine_Provider_AD_%20Risk_Discussion_Guide.pdf [accessed 2021-01-13]
9. Lader M. Benzodiazepines revisited--will we ever learn? Addiction 2011 Dec;106(12):2086-2109. [doi: 10.1111/j.1360-0443.2011.03563.x] [Medline: 21714826]
10. Tannenbaum C, Martin P, Tamblyn R, Benedetti A, Ahmed S. Reduction of inappropriate benzodiazepine prescriptions among older adults through direct patient education: the EMPOWER cluster randomized trial. JAMA Intern Med 2014 Jun 01;174(6):890-898. [doi: 10.1001/jamainternmed.2014.949] [Medline: 24733354]
11. Martin P, Tamblyn R, Ahmed S, Tannenbaum C. An educational intervention to reduce the use of potentially inappropriate medications among older adults (EMPOWER study): protocol for a cluster randomized trial. Trials 2013 Mar 20;14(1):80 [FREE Full text] [doi: 10.1186/1745-6215-14-80] [Medline: 23514019]
12. Pharmacy benefits management services. U.S. Department of Veterans Affairs. URL: https://www.pbm.va.gov/ [accessed 2022-08-28]
13. Park TW, Saitz R, Ganoczy D, Ilgen MA, Bohnert AS. Benzodiazepine prescribing patterns and deaths from drug overdose among US veterans receiving opioid analgesics: case-cohort study. BMJ 2015 Jun 10;350(jun10 9):h2698 [FREE Full text] [doi: 10.1136/bmj.h2698] [Medline: 26063215]
14. Gosselin P, Ladouceur R, Morin CM, Dugas MJ, Baillargeon L. Benzodiazepine discontinuation among adults with GAD: a randomized trial of cognitive-behavioral therapy. J Consult Clin Psychol 2006 Oct;74(5):908-919. [doi: 10.1037/0022-006x.74.5.908]

15. Gould RL, Coulson MC, Patel N, Highton-Williamson E, Howard RJ. Interventions for reducing benzodiazepine use in older people: meta-analysis of randomised controlled trials. Br J Psychiatry 2014 Feb 02;204(2):98-107. [doi: 10.1192/bjp.bp.113.126003] [Medline: 24493654]
16. Vicens C, Bejarano F, Sempere E, Mateu C, Fiol F, Socias I, et al. Comparative efficacy of two interventions to discontinue long-term benzodiazepine use: cluster randomised controlled trial in primary care. Br J Psychiatry 2014 Jun 02;204(6):471-479. [doi: 10.1192/bjp.bp.113.134650] [Medline: 24526745]
17. Meffert BN, Morabito DM, Sawicki DA, Hausman C, Southwick SM, Pietrzak RH, et al. US veterans who do and do not utilize veterans affairs health care services. Prim Care Companion CNS Disord 2019 Jan 17;21(1):18m02350. [doi: 10.4088/pcc.18m02350]
18. Hadlandsmyth K, Mosher H, Vander Weg MW, Lund BC. Decline in prescription opioids attributable to decreases in long-term use: a retrospective study in the veterans health administration 2010-2016. J Gen Intern Med 2018 Jun 29;33(6):818-824 [FREE Full text] [doi: 10.1007/s11606-017-4283-8] [Medline: 29380212]
19. What is the ADDIE Model? International Society for Educational Technology. URL: https://www.isfet.org/pages/addie-model [accessed 2021-12-02]
20. Sung S, Park H. Effect of a mobile app-based cultural competence training program for nurses: a pre- and posttest design. Nurse Educ Today 2021 Apr;99:104795. [doi: 10.1016/j.nedt.2021.104795] [Medline: 33621852]
21. Dodgson JE. Reflexivity in qualitative research. J Hum Lact 2019 May 08;35(2):220-222. [doi: 10.1177/0890334419830990] [Medline: 30849272]
22. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
23. Sobo EJ, Seid M, Reyes Gelhard L. Parent-identified barriers to pediatric health care: a process-oriented model. Health Serv Res 2006 Feb;41(1):148-172 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00455.x] [Medline: 16430605]
24. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516 [FREE Full text] [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
25. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
26. Artieta-Pinedo I, Paz-Pascual C, Bully P, Espinosa M, EmaQ Group. Design of the maternal website EMAeHealth that supports decision-making during pregnancy and in the postpartum period: collaborative action research study. JMIR Form Res 2021 Aug 09;5(8):e28855 [FREE Full text] [doi: 10.2196/28855] [Medline: 34383670]
27. Neville C, Da Costa D, Rochon M, Peschken CA, Pineau CA, Bernatsky S, et al. Development of the lupus interactive navigator as an empowering web-based eHealth tool to facilitate lupus management: users perspectives on usability and acceptability. JMIR Res Protoc 2016 May 30;5(2):e44 [FREE Full text] [doi: 10.2196/resprot.4219] [Medline: 27240666]
28. Sweeney L, Windgassen S, Artom M, Norton C, Fawson S, Moss-Morris R. A novel digital self-management intervention for symptoms of fatigue, pain, and urgency in inflammatory bowel disease: describing the process of development. JMIR Form Res 2022 May 18;6(5):e33001 [FREE Full text] [doi: 10.2196/33001] [Medline: 35583924]
29. Stetler CB, Legro MW, Wallace CM, Bowman C, Guihan M, Hagedorn H, et al. The role of formative evaluation in implementation research and the QUERI experience. J Gen Intern Med 2006 Feb;21 Suppl 2:S1-S8 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00355.x] [Medline: 16637954]
30. Patel SR, Margolies PJ, Covell NH, Lipscomb C, Dixon LB. Using instructional design, analyze, design, develop, implement, and evaluate, to develop e-learning modules to disseminate supported employment for community behavioral health treatment programs in New York state. Front Public Health 2018 May 7;6:113 [FREE Full text] [doi: 10.3389/fpubh.2018.00113] [Medline: 29868533]
31. Srikesavan CS, Williamson E, Eldridge L, Heine P, Adams J, Cranston T, et al. A web-based training resource for therapists to deliver an evidence-based exercise program for rheumatoid arthritis of the hand (iSARAH): design, development, and usability testing. J Med Internet Res 2017 Dec 13;19(12):e411 [FREE Full text] [doi: 10.2196/jmir.8424] [Medline: 29237581]
32. Babatunde FO, MacDermid J, Grewal R, Macedo L, Szekeres M. Development and usability testing of a web-based and therapist-assisted coping skills program for managing psychosocial problems in individuals with hand and upper limb injuries: mixed methods study. JMIR Hum Factors 2020 May 06;7(2):e17088 [FREE Full text] [doi: 10.2196/17088] [Medline: 32374265]

---

## Abbreviations

**ADDIE:** Analysis, Design, Development, Implementation, and Evaluation
**CFIR:** Consolidated Framework for Implementation Research
**EMPOWER:** Eliminating Medications Through Patient Ownership of End Results
**EMPOWER-ED:** Eliminating Medications Through Patient Ownership of End Results Electronic Delivery
**PI:** principal investigator

**VHA:** Veterans Health Administration

*Edited by T Leung; submitted 08.12.21; peer-reviewed by F Babatunde, S Stones; comments to author 29.06.22; revised version received 21.07.22; accepted 10.08.22; published 19.09.22.*

*Please cite as:*
*Cucciare MA, Abraham TH, Kemp L, White P, Marchant K, Hagedorn HJ, Humphreys K*
*Adapting the Eliminating Medications Through Patient Ownership of End Results Protocol to Promote Benzodiazepine Cessation Among US Military Veterans: Focus Group Study With US Military Veterans and National Veterans Health Administration Leaders*
*J Med Internet Res 2022;24(9):e35514*
URL: *https://www.jmir.org/2022/9/e35514*
doi:*10.2196/35514*
PMID:*36121697*

Original Paper

# Preliminary Investigation of Shift, a Novel Smartphone App to Support Junior Doctors' Mental Health and Well-being: Examination of Symptom Progression, Usability, and Acceptability After 1 Month of Use

Samineh Sanatkar[1,2], PhD; Isabelle Counson[1,2], GradDipPsychAdv, MPH; Andrew Mackinnon[1], PhD; Alexandra Bartholomew[1], BCom (BSc), MPH; Nick Glozier[3], MBBS, FRANZCP, PhD; Samuel Harvey[1,2], MBBS, MRCGP, MRCPsych, FRANZCP, PhD

[1]Black Dog Institute, Randwick, Australia
[2]School of Psychiatry, Faculty of Medicine, University of New South Wales Sydney, Kensington, Australia
[3]Central Clinical School, Faculty of Medicine and Health, University of Sydney, Camperdown, Australia

**Corresponding Author:**
Samineh Sanatkar, PhD
Black Dog Institute
Hospital Road
Randwick, 2031
Australia
Phone: 61 02 9065 9179
Email: s.sanatkar@unsw.edu.au

**Related Article:**

This is a corrected version. See correction statement: https://www.jmir.org/2023/1/e47277

## Abstract

**Background:** *Shift* is a novel smartphone app for providing a digital-first mental health resource to junior doctors. It contains psychoeducational material, cognitive behavioral modules, guided mediations, information on common work stressors, and a section on help-seeking options for psychological problems through workplace and private avenues.

**Objective:** This study aimed to conduct a preliminary investigation of the use and potential effectiveness of *Shift* on depressive and anxiety symptoms (primary outcomes) and work and social functioning, COVID-19 safety concerns, and help seeking (secondary outcomes). This study also sought feedback on whether *Shift* was seen as an acceptable tool.

**Methods:** Junior doctors in New South Wales, Australia, were approached through promotional activities from the Ministry of Health, specialist medical colleges, and social media advertisements between June and August 2020. Consenting participants provided web-based baseline data, used the *Shift* app for 30 days, and were asked to complete a poststudy web-based questionnaire. Outcomes were analyzed under the intention-to-treat principle.

**Results:** A total of 222 (n=156 female, 70.3%; mean age 29.2, SD 4.61 years) junior doctors provided full baseline data. Of these, 89.2% (198/222) downloaded the app, logged into the app approximately 6 times (mean 5.68, SD 7.51), completed 4 in-app activities (mean 3.77, SD 4.36), and spent a total of 1 hour on in-app activities (mean 52:23, SD 6:00:18) over 30 days. Postintervention and app use data were provided by 24.3% (54/222) of participants. Depressive and anxiety symptoms significantly decreased between the pre- and postassessment points as expected; however, physicians' COVID-19 safety concerns significantly increased. Work and social functioning, COVID-19 concerns for family and friends, and help seeking did not change significantly. There was no significant relationship between symptom changes and app use (number of log-ins, days between first and last log-in, and total activity time). Most poststudy completers (31/54, 57%) rated *Shift* highly or very highly.

**Conclusions:** Despite high levels of nonresponse to the poststudy assessment and increases in COVID-19 safety concerns, junior doctors who used the app reported some improvements in depression and anxiety, which warrant further exploration in a robust manner.

*(J Med Internet Res 2022;24(9):e38497)* doi:10.2196/38497

**KEYWORDS**

digital mental health; mobile health apps; mHealth apps; help seeking; junior doctors; depression; mobile phone

## Introduction

### Background

Psychological distress and mental health disorders such as major depression are prevalent in junior doctors [1,2]. A recent review and meta-analysis identified high work demands, patient care concerns, and a poor work environment as the most important workplace risk factors associated with the decline of junior doctors' mental health [1]. Individual risk factors such as low perceived mental and physical well-being and poor self-efficacy were also identified, although they were generally less robust predictors of distress [1]. Poor mental health can negatively affect junior doctors' prospects at a time that may be critical for establishing their personal life and career goals. For example, poor perceived mental and physical health among doctors is associated with lower job satisfaction and intentions to leave medicine altogether [3]. Particularly among junior doctors, job burnout is further associated with decreased professionalism and lower patient safety [4]. Thus, when severe distress impacts workplace functioning, the quality of care provided to patients and early-career physicians' long-term job prospects may be negatively impacted [5,6].

National survey data indicated that 21% of surveyed Australian doctors had been diagnosed with major depression, and 9% had received an anxiety disorder diagnosis in their lifetime, with junior doctors reporting particularly high psychological distress at the time of survey completion [7]. A recent review and meta-analysis suggested that psychological interventions aimed at reducing common mental health disorders in physicians lessen disease burden [8]. Although this evidence is promising, the authors also noted that the delivery and acquisition of cognitive behavioral and mindfulness techniques requires doctors to set aside time and resources to pursue such therapy [8]. This could constitute a substantial barrier to entering treatment. The authors suggested that it is therefore advised to explore the option of delivering therapeutic components on the web or via smartphone apps to increase accessibility [8]. To our knowledge, no attempt has been made to date to convey the principles of cognitive behavioral therapy (CBT) and mindfulness to populations via a smartphone app.

Australian junior doctors seek psychological treatment for depression at lower rates than senior doctors [9]. In addition to issues related to mental health stigma in the medical profession [10,11], one practical reason for this may be a gap in the knowledge on how to seek such support. A recent large-scale Australian survey indicated that one-quarter of surveyed junior doctors (1929/7715, 25%) did not know or were unsure of how to seek help for their physical and mental health concerns [12]. Although some mental health problems remit over time [13], prognosis is generally best if individuals seek treatment early on [14,15]. Untreated persistence of mental health problems over prolonged periods reduces recovery rates, even when therapy is eventually sought [15]. Thus, facilitating access to professional support early on may be an important factor in increasing the success rate of interventions aimed at improving junior doctors' mental health.

Several prominent mental health crises in junior physician populations have been documented in the media in recent years [16,17]. In the Australian context, a cluster of 3 junior doctors' suicides in New South Wales gained widespread public interest, and calls for reforms were made from inside and outside the medical profession [16,18]. As a response to these calls and after consultation with relevant parties, including junior doctors, the New South Wales Ministry of Health issued a 10-point Junior Medical Officer Wellbeing and Support Plan [19]. One of the 10 initiatives was the development and implementation of the *Shift* smartphone app. *Shift* was developed to provide a discreet resource for junior doctors in New South Wales to access mental health–related content and information on how to seek mental health care.

*Shift* digitally delivers therapeutic components adapted from CBT, acceptance and commitment therapy, mindfulness, and psychoeducational material shown to be useful in physician and health care worker populations [8,20,21] with the primary aim of reducing the symptoms of depression and anxiety. The current version of *Shift* is a result of an iterative development process involving junior doctors in qualitative interviews, user experience workshops, and pilot testing of the prototype version of the app. The full development and testing process is outlined in a previous publication [22]. The COVID-19 outbreak coincided with the recruitment phase of this study. To respond to the new situation, 2 clinical psychologists and a psychiatrist at the Black Dog Institute (SH) helped develop novel contents on health anxiety and COVID-19 concerns, the latter of which was prominently placed within the app.

### Objective

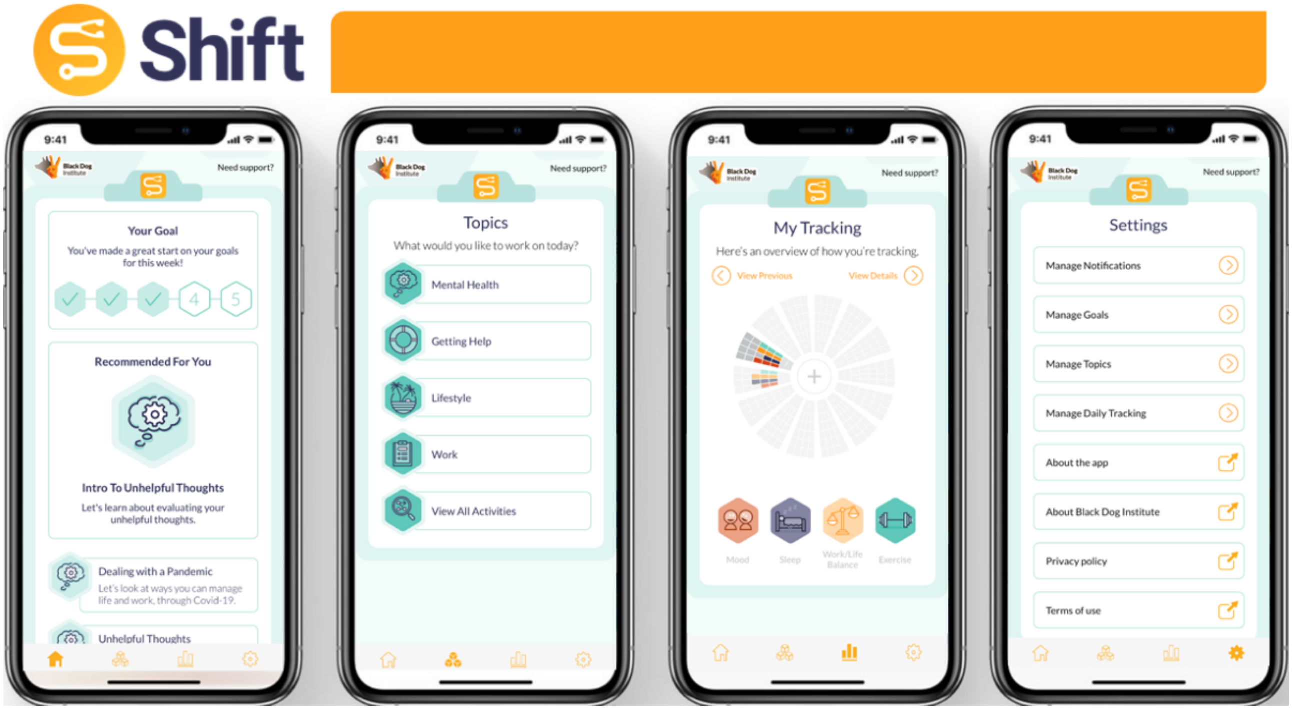
This study sought preliminary evidence of the effectiveness of the *Shift* app among junior doctors. This study was planned as a 2-arm waitlist-controlled randomized controlled trial, but as a response to the COVID-19 pandemic and to ensure that all junior doctors interested in taking part in this investigation could download the app without delay, this study was implemented as an uncontrolled quasi-experimental pre-post intervention evaluation with the primary aim of assessing depressive and anxiety symptom levels over the course of 1 month of using the app. Relatedly, the study examined whether potential improvements in depressive symptoms would help restore functioning across work and private life domains and whether COVID-19 pandemic concerns remitted over the course of the study. Where significant improvements were observed, we aimed at investigating whether there was a relationship between app use and symptom reduction. Another aim was to encourage help seeking for mental health problems by way of providing in-app information on the various workplace and private avenues of available support and by alleviating potential concerns around mandatory reporting guidelines.

## Methods

### Study Design

We conducted an uncontrolled quasi-experimental pre-post intervention to examine the trajectories of depression and anxiety symptoms (primary outcomes) and COVID-19 pandemic concerns, work and social functioning, and help-seeking tendencies (secondary outcomes). This information was provided via web-based questionnaires. *Shift* app use data were automatically collected in the 1-month period between the participants' download of *Shift* and the poststudy survey assessment. An overview of the app layout design can be seen in Figure 1. At the end of the study period, we reassessed the participants' symptom scores and asked the participants to provide app usability and acceptability ratings using web-based questionnaires.

**Figure 1.** Examples of *Shift* app screen designs.



You've made a great start on your goals for this week!

Let's learn about evaluating your unhelpful thoughts.

Let's look at ways you can manage life and work, through Covid-19.

What would you like to work on today?

Here's an overview of how you're tracking.

Mixed model repeated measures ANOVA was used to investigate pre- and postsymptom change in the primary and secondary outcome variables. This method was used in preference to paired sample *t* tests to retain all observations. It also makes less stringent assumptions regarding missingness mechanisms. We further explored whether app use was indicative of significant symptom trajectories. The relation between app use and poststudy symptom levels was examined using linear regression analyses, whereby postintervention symptom scores were the dependent variable and app use variables were entered as independent variables along with baseline symptom scores as a covariate.

### Participants

Between June and August 2020, junior medical officers practicing in New South Wales, Australia, were invited to join the *Shift* app research study. Recruitment took place through three different avenues: (1) the New South Wales Ministry of Health promotional efforts through an official media release and at relevant meetings or forums, (2) liaison with medical colleges and other organizations to disseminate study recruitment material, and (3) through a social media campaign via Facebook and on the Black Dog Institute website. Disseminated recruitment materials contained study details (eg, study components, app contents, and study duration), ethical information (eg, voluntary participation and data handling), and a link to a purpose-designed study website that provided the full Participant Information Statement and Consent Form as a downloadable file.

Eligibility criteria were currently being employed as an intern, a resident, a registrar, or a junior career medical officer in New South Wales and owning an internet-enabled smartphone with an iOS or Android operating system. After consenting and confirming eligibility through screening, the participants completed a web-based questionnaire containing demographic information and mental health–related questions. Subsequently, the participants were given instructions on how to download the *Shift* app from the App Store or Google Play Store. The onboarding of the *Shift* app was completed by the participants, which included prompts to set preferences with regard to whether to allow app notifications, and if yes, at which intervals, and to set weekly goals for app activity completion. These settings could be modified within the app at any time. After 30 days, the participants were asked to complete a postintervention web-based assessment questionnaire. Items reassessed the participants' mental health status and asked for *Shift* app feedback. Measurement tools are provided in the subsequent sections.

## Measures

### Depression

The participants completed the 9-item Patient Health Questionnaire [23], which assesses major depression symptom severity. The participants responded to items such as "feeling down, depressed, or hopeless" on a 4-point Likert-type scale ranging from 0 (*not at all*) to 3 (*nearly every day*). Total scores range from 0 to 27, with scores of ≥10 being indicative of possible major depressive disorder. The measure has been shown to be reliable and valid and has been used in physician samples, including training physicians [23,24].

### Anxiety

General anxiety symptom levels were measured using the 7-item Generalized Anxiety Disorder scale [25]. The participants indicated their agreement with items such as "feeling nervous, anxious or on edge" on a Likert-type scale ranging from 0 (*not at all*) to 3 (*nearly every day*). Total scores on the 7-item Generalized Anxiety Disorder scale range from 0 to 21, with scores ≥10 being indicative of symptoms consistent with general anxiety disorder. The measure has been found to have good psychometric properties and has been used in physician samples, including during the COVID-19 pandemic [25-27].

### Work and Social Functioning

Functioning in social and work arenas was measured using the 5-item Work and Social Adjustment Scale [28]. The participants responded to items such as "Because of my problems, my ability to work is impaired" on an 8-point Likert-type scale anchored at 0 (*not at all*) and 8 (*very severely*). Scores on the Work and Social Adjustment Scale range from 0 to 40, with higher scores indicating reductions in work and social functioning. Specifically, scores between 10 and 20 are indicative of significant functional impairment, and scores >20 indicate functional impairment on clinical symptom levels.

### COVID-19 Safety Concerns

The participants' safety concerns regarding the possible contraction of COVID-19 were assessed using 2 items developed for the study. The participants indicated their concerns for their own safety ("How concerned or worried are you that you, personally, will catch COVID-19?") and the safety of their family and friends ("How concerned or worried are you that your family members or friends will catch COVID-19?") on a 5-point Likert-type scale ranging from 1 (*not at all*) to 5 (*extremely*).

### Help Seeking

To assess the participants' willingness to seek help for mental health concerns, we modified the first item of the General Help-Seeking Questionnaire [29]. The original item "If you were having a personal or emotional problem, how likely is it that you would seek help from the following people?" was amended to "If you were having a personal or emotional problem, how likely is it that you would seek help from a mental health professional?" The participants indicated their responses on a 7-point Likert-type scale ranging from 1 (*extremely unlikely*) to 7 (*extremely likely*). To assess recent actual help-seeking behaviors for mental health concerns, we created the following single-item indicator: "Did you seek any help for a personal or emotional problem from a mental health professional in the past month?" The participants responded to this item using a "yes" or "no" response option.

### App Acceptability

The poststudy questionnaire included several *Shift* app acceptability and usability items. Acceptability items were as follows: "What is your overall rating of the app?" "I would use this app again in the future." "I would recommend this app to other Doctors in Training." The participants indicated their overall rating of the app on a 5-point Likert-type scale ranging from 1 (*very poor*) to 5 (*very high*) and stated their agreement with the future app use and recommendation items on a 5-point Likert-type response scale anchored at 1 (*strongly disagree*) and 5 (*strongly agree*). For a brief indication of the perceived usability of the *Shift* app, the participants were asked to respond to a modified single-item indicator adopted from the System Usability Scale [30], termed "I thought the app was easy to use." The participants indicated their level of agreement with this item on a 5-point Likert-type response scale anchored at 1 (*strongly disagree*) and 5 (*strongly agree*). As a part of the app onboarding process, the participants set their own app use goals (the app recommended the completion of between 3 and 5 activities per week). The participants were asked whether they reached their own app use goals: "On average, did you hit your weekly challenge goal or not?" to which the participants indicated 1 (*hit goal*) or 0 (*missed goal*).

### App Use Variables

*Shift* app use measurement included the total number of log-ins, days between the first and last log-in (maximum 30 days), and the number of activities started and completed. Examples for cognitive behavioral activities were "unhelpful thoughts" and "evaluating thoughts." "Sleep health" and "adjusting to shift work" were examples for work and lifestyle activities. A full list of activities and how they were organized within the app can be found in Multimedia Appendix 1. Total activity time and times spent on each of the cognitive behavioral, work and lifestyle, mindfulness, psychoeducation, value-based actions, and "get help" sections were also recorded, although it was not possible to distinguish between the times when users engaged with the app contents and the times when the app was open but not attended to.

## Ethics Approval

The study was prospectively registered with the Australian New Zealand Clinical Trials Registry under the trial number ACTRN12620000571976. Ethics approval for this study was obtained through University of New South Wales Sydney Human Research Ethics Committee (protocol number HC200212).

# Results

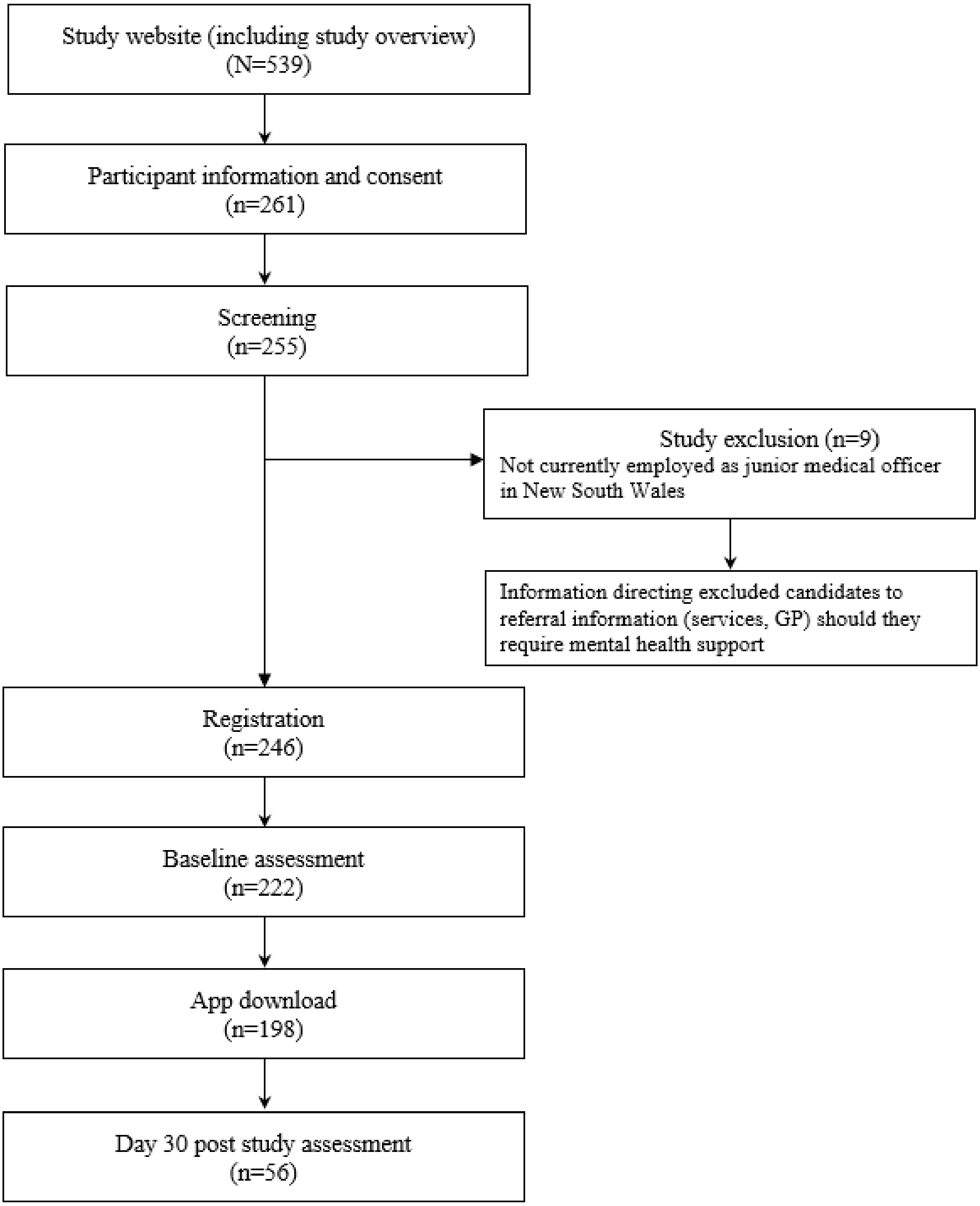
## Sample Characteristics

### Participant Flow

Figure 2 illustrates the participant flow through the *Shift* study. Of the 539 doctors who visited the study website, 261 (48.4%)

consented to participate and 255 (47.3%) completed the study screen. Of these 255 candidates, 9 (3.5%) were ineligible to register for the study because they were not currently employed as junior doctors in New South Wales, and 24 (9.4%) candidates did not complete the baseline assessment, resulting in a final sample size of 222 study participants. Although all but 1.8% (4/222) of participants went on to download the *Shift* app as directed, only 68.9% (153/222) accessed the *Shift* app more than once. Attrition in the poststudy assessment was high, with 25.2% (56/222) of the participants, that is, a quarter of baseline completers, returning the full poststudy assessment after 1 month. A total of 4% (2/56) of participants returned the baseline and poststudy assessments but did not download the app. Consequently, complete data were available for 24.3% (54/222) of the participants.

**Figure 2.** Flowchart illustrating how participants progressed through the Shift study. GP: general practitioner.



### *Baseline Sample Composition*

The participants were predominantly female (156/222, 70.3%) and had a mean age of 29.2 (SD 4.61) years. Most participants were employed as registrars (85/222, 38.3%), interns (55/222, 24.8%), or residents (50/222, 22.5%) in metropolitan areas (137/222, 61.7%). Symptoms of poor mental health were noted among some participants; 23.9% (53/222) reported depressive symptoms consistent with possible major depression, 14.4% (32/222) reported symptoms indicative of general anxiety disorder, and 9.5% (21/222) indicated clinically significant levels of functional impairment. In the month preceding the baseline assessment, 19.5% (43/222) had sought professional help for mental health concerns, and 32.9% (73/222) of doctors indicated having had contact with a patient with observed or covert COVID-19.

Table 1 shows the baseline characteristics of participants who completed only the baseline assessment (166/222, 74.8%) and those who completed the baseline and also poststudy assessments (56/222, 25.2%). Comparisons of demographic (age and gender), mental health, functioning, COVID-19 concerns, and help-seeking information at baseline did not suggest any systematic differences between the study completers and those who did not complete the poststudy assessment (all *P*>.05; Table 1). The standardized mean differences between those who completed both the pre- and postassessments and those who were lost to follow-up were small, with corrected effect size values (Hedges *g*) ranging from 0.08 for personal COVID-19 safety concerns to 0.29 for work and social functioning.

**Table 1.** Sample characteristics at baseline (N=222), comparing participants who completed the prestudy assessment (n=166) and those who completed pre- and poststudy assessments (n=56).

| Characteristics | Prestudy assessment only | Pre- and poststudy assessment | *F*/$\chi^2$ (*df*) | *P* value | Hedges g/Cohen *w* |
|---|---|---|---|---|---|
| **Demographics** | | | | | |
| Age (years), mean (SD) | 28.93 (4.61) | 29.8 (4.6) | 1.49 (1) | .22 | 0.19 |
| Female, n (%) | 118 (71.1) | 38 (67.9) | 1.89 (2) | .39 | 0.09 |
| **Level of training, n (%)** | | | 4.67 (4) | .32 | 0.10 |
| Registrar | 58 (34.9) | 27 (48.2) | | | |
| Intern | 45 (27.1) | 10 (17.9) | | | |
| Resident | 40 (24.1) | 10 (17.9) | | | |
| Senior resident | 22 (13.3) | 9 (16.1) | | | |
| Junior career medical officer | 1 (0.6) | 0 (0) | | | |
| **Specialty[a], n (%)** | | | 26.10 (21) | .20 | 0.11 |
| General medicine | 11 (6.6) | 3 (5.4) | | | |
| Psychiatry | 10 (6.0) | 2 (3.6) | | | |
| Emergency medicine | 8 (4.8) | 5 (8.9) | | | |
| Pediatric and child health | 7 (4.2) | 5 (8.9) | | | |
| **Work location, n (%)** | | | 1.54 (2) | .46 | 0.08 |
| Metropolitan | 99 (59.6) | 38 (67.9) | | | |
| Regional | 53 (31.9) | 13 (23.2) | | | |
| Rural | 14 (8.4) | 5 (8.9) | | | |
| **Mental health, mean (SD)** | | | | | |
| Depression | 6.99 (4.12) | 6.39 (5.03) | 0.78 (1) | .38 | 0.14 |
| Anxiety | 5.45 (3.94) | 5.07 (4.39) | 0.36 (1) | .55 | 0.09 |
| Work and social adjustment | 10.92 (7.32) | 8.84 (6.77) | 3.49 (1) | .06 | 0.29 |
| **COVID-19 pandemic concerns, mean (SD)** | | | | | |
| Self | 1.90 (0.77) | 1.84 (0.68) | 0.26 (1) | .61 | 0.08 |
| Family and friends | 2.57 (1.01) | 2.48 (0.95) | 0.30 (1) | .59 | 0.09 |
| **Help seeking** | | | | | |
| Willingness to seek help, mean (SD) | 4.14 (1.41) | 3.91 (1.24) | 1.15 (1) | .28 | 0.17 |
| Sought help in past month, n (%) | 33 (19.9) | 10 (17.9) | 0.11 (1) | .74 | 0.02 |

[a] Only the most common specialty areas are listed out of a total of 22.

## Primary Outcomes

### *Depression Symptom Changes*

As shown in Table 2, depression symptom scores significantly decreased between the pre- (estimated marginal mean [EMM] 6.84, SE 0.29) and poststudy (EMM 5.24, SE 0.38) assessments, with $t_{74.89}=4.34$ and $P<.001$. Glass delta effect sizes for depression symptoms changes (using the SD at baseline) were small to medium ($\Delta=0.37$).

**Table 2.** Linear mixed effect models to examine changes in responses between pre- and poststudy assessments.

| Variable | Prestudy EMM[a] (SE) | Poststudy EMM (SE) | *t* test (*df*) | *P* value | Δ[b] |
|---|---|---|---|---|---|
| **Depression** | 6.84 (0.29) | .24 (0.38) | | | |
| Intercept | | | 13.94 (80.25) | <.001 | N/A[c] |
| Time | | | 4.34 (74.89) | <.001 | .37 |
| **Anxiety** | 5.35 (0.27) | 4.44 (0.42) | | | |
| Intercept | | | 10.52 (70.65) | <.001 | N/A |
| Time | | | 2.20 (67.16) | .03 | .22 |
| **Functioning** | 10.40 (0.49) | 9.77 (0.78) | | | |
| Intercept | | | 12.51 (71.16) | <.001 | N/A |
| Time | | | 0.84 (62.42) | .41 | .09 |
| **COVID-19 pandemic concerns, self** | 1.88 (0.05) | 2.05 (0.09) | | | |
| Intercept | | | 25.47 (70.33) | <.001 | N/A |
| Time | | | −2.16 (64.77) | .04 | .23 |
| **COVID-19 pandemic concerns, family** | 2.55 (0.07) | 2.60 (0.10) | | | |
| Intercept | | | 25.66 (71.21) | <.001 | N/A |
| Time | | | −0.60 (66.10) | .55 | .06 |
| **Help-seeking willingness** | 4.08 (0.09) | 4.03 (0.15) | | | |
| Intercept | | | 26.57 (72.61) | <.001 | N/A |
| Time | | | 0.35 (60.77) | .73 | .04 |
| **Help seeking in the past month** | .19 (0.03) | .13 (0.04) | | | |
| Intercept | | | 3.10 (61.17) | .003 | N/A |
| Time | | | 1.50 (72.82) | .14 | .17 |

[a] EMM: estimated marginal mean.

[b] Δ: Glass delta effect size.

[c] N/A: not applicable.

### *Anxiety Symptom Changes*

Results from linear mixed effects analysis (Table 2) indicated that anxiety symptoms significantly decreased between the pre- (EMM 5.35, SE 0.27) and poststudy (EMM 4.44, SE 0.42) assessments, with $t_{67.16}=2.20$ and $P=.032$. Glass delta effect sizes for anxiety symptom changes were small ($\Delta=0.22$).

## Secondary Outcomes

### *Functioning, COVID-19 Pandemic Concerns, and Help-Seeking Changes*

Table 2 further summarizes the secondary outcomes pertaining to expected improvements in work and social functioning, personal safety concerns, concerns regarding family and friends contracting COVID-19, willingness to seek help, and recent help-seeking behaviors. Although work and social functioning scores improved as expected, changes between the pre- and poststudy assessment points were not significant ($P>.05$). Pandemic-related concerns toward participants’ own safety significantly increased for poststudy completers (EMM 2.05, SE 0.09) compared with baseline reports (EMM 1.88, SE 0.05; $t_{64.77}=-2.16$; $P=.04$; $\Delta=0.23$). Concerns for family members and friends as well as help-seeking intentions largely remained unchanged and were not significant.

### *App Use*

Over the 30-day study period, the participants logged in an average of 6 times (mean 5.68, SD 7.51) and used the app over a period of approximately 2 weeks (mean 14.50, SD 12.78). On average, the participants started approximately 5 activities (mean 5.44, SD 5.94) during this time and completed 4 of them (mean 3.77, SD 4.36). Overall, the participants spent just under an hour on the activities presented within the app (mean 52:23, SD 6:00:18). CBT module use constituted the most used component with an average time spent on CBT activities of about 12 minutes (mean 12:05, SD 1:18:42), followed by about 4 minutes on work and lifestyle activities (mean 03:55, SD 10:39) and 3 minutes on mindfulness (mean 03:05, SD 11:54). On average, the participants spent the least time viewing psychoeducation (mean 00:22, SD 01:59), values (mean 00:11, SD 01:12) and get help (mean 00:04, SD 00:34) activities. Owing to the large range of times recorded, we assumed that some participants had their app open without actively engaging with the contents, prompting us to remove outliers. Although outliers >3 SDs from the mean were removed, the variability of app use data remained large.

App use behavior differed substantially between study completers and those lost to follow-up, as shown in Table 3. App use means of study completers for log-in, days between first and last log-in, and app activity use were approximately 1 SD above those who did not return the poststudy assessment (Hedges *g* ≥1.0), and the overall time spent on activities was over half an SD greater among study completers (Hedges *g*=0.61). Although there seemed to be no significant differences with regard to baseline characteristics, participants who completed the poststudy questionnaire were more involved in the month-long app use component of this research than those who did not complete the poststudy questionnaire.

**Table 3.** Participants' 30-day *Shift* app use and poststudy app acceptability ratings, comparing participants who downloaded *Shift* and completed only the prestudy assessment (n=144) and those who downloaded *Shift* and completed pre- and poststudy assessments (n=54)[a].

| Variable | Prestudy assessment only, mean (SD) | Pre- and poststudy assessment, mean (SD) | *F* (*df*) | *P* value | Hedges *g* |
|---|---|---|---|---|---|
| Log-ins | 3.56 (3.10) | 11.45 (11.83) | 54.50 (1,195) | <.001 | 1.18 |
| Days between first and last log-in | 11.26 (11.82) | 23.09 (11.16) | 40.46 (1,195) | <.001 | 1.02 |
| Activities started | 4.01 (4.91) | 9.45 (6.74) | 37.62 (1,196) | <.001 | 1.00 |
| Activities completed | 2.68 (3.54) | 6.88 (4.97) | 41.74 (1,196) | <.001 | 1.06 |
| Total activity time[b] | 12:23 (27:42) | 01:08:13 (02:51:22) | 14.30 (1,195) | <.001 | 0.61 |
| Time spent on CBT[c] | 05:35 (20:09) | 29:17 (02:26:18) | 3.60 (1,196) | .06 | 0.30 |
| Time spent on work and lifestyle | 02:22 (06:39) | 08:13 (16:52) | 12.14 (1,194) | .01 | 0.56 |
| Time spent on mindfulness | 02:24 (12:50) | 04:57 (08:43) | 1.80 (1,196) | .18 | 0.21 |
| Time spent on psychoeducation | 00:12 (01:07) | 00:48 (03:21) | 3.48 (1,193) | .06 | 0.30 |
| Time spent on values | 00:09 (01:10) | 00:17 (01:18) | 0.37 (1,189) | .55 | 0.11 |
| Time spent on get help | 00:03 (00:34) | 00:08 (00:35) | 0.76 (1,196) | .38 | 0.15 |

[a] Outliers >3 SDs from the mean were removed from use data.

[b] Activity time is given in minutes and seconds (00:00) or hours, minutes, and seconds (00:00:00).

[c] CBT: cognitive behavioral therapy.

### App Use Mental Health Trajectory Relation

Given the sample size, 1 use predictor was examined at a time. Where present, outliers 3 SDs from the mean were removed for these analyses because we assumed that very long use times indicated that the app was open but not actively used. Very high correlations among the number of activities started, activities completed, and time spent on activities (*r* ranged between 0.83 and 0.96) indicated that these variables measured a singular construct (ie, activity engagement). To avoid redundancy in use variables, we only entered the time spent on activities, number of log-ins, and days between participants' first and last log-in. As indicated in Table 4, none of the app use measures predicted depression and anxiety outcomes or COVID-19 pandemic safety concerns (*P*>.05). Although the effects for depression and anxiety were in the expected direction (eg, each new log-in predicted about 0.7 of a point reduction on the 9-item Patient Health Questionnaire), no significant relation between app use and symptom change could be detected in the present sample with the inclusion of log-in and activity use count data.

**Table 4.** Linear regression analyses of the relation between app use variables and changes in depression and anxiety symptoms and junior doctors' COVID-19 safety concerns, with baseline values entered as covariates (N=54).

| Variable | Constant, b (SE-b) | Baseline, b (SE-b) | b | SE-b | β | $R^2$ value | Δ $R^2$ value | Δ *F* (*df*) | *P* value |
|---|---|---|---|---|---|---|---|---|---|
| **Depression** | | | | | | | | | |
| Log-ins | 2.14 (1.13) | 0.52 (0.07) | −0.65 | 1.00 | −0.06 | 0.52 | 0.00 | 0.42 (1,50) | .52 |
| Days between first and last log-in | 1.64 (1.14) | 0.53 (0.07) | −0.20 | 0.82 | −0.02 | 0.51 | 0.00 | 0.06 (1,51) | .81 |
| Total activity time (min) | 11.89 (0.86) | 0.53 (0.07) | −0.40 | 0.52 | −0.08 | 0.52 | 0.01 | 0.59 (1,51) | .45 |
| **Anxiety** | | | | | | | | | |
| Log-ins | 1.80 (1.30) | 0.60 (0.10) | −0.44 | 1.15 | −0.04 | 0.43 | 0.00 | 0.15 (1,50) | .70 |
| Days between first to last log-in | 2.08 (1.24) | 0.61 (0.10) | −0.61 | 0.93 | −0.07 | 0.44 | 0.01 | 0.43 (1,51) | .51 |
| Total activity time (min) | 11.92 (0.97) | 0.60 (0.10) | −0.43 | 0.59 | −0.08 | 0.44 | 0.01 | 0.54 (1,51) | .47 |
| **COVID-19 pandemic concerns, self** | | | | | | | | | |
| Log-ins | 11.00 (0.28) | 0.48 (0.12) | .15 | 0.22 | .08 | 0.27 | 0.01 | 0.44 (1,50) | .51 |
| Days between first to last log-in | 0.92 (0.29) | 0.47 (0.12) | .18 | 0.18 | .49 | 0.27 | 0.01 | 0.99 (1,51) | .32 |
| Total activity time (minutes) | 11.02 (0.25) | 0.47 (0.12) | .11 | 0.12 | .12 | 0.27 | 0.01 | 0.92 (1,51) | .34 |

### *App Acceptability Ratings*

The participants completed app acceptability items as part of the poststudy survey assessment. All average ratings were above the midpoint of the response scales. Endorsement of the ease-of-use item was good (mean 4.09, SD 0.73) as well as the participants' willingness to use the app again in the future (mean 3.67, SD 0.97) and their willingness to recommend the app to colleagues (mean 3.70, SD 0.92). The overall rating of the app was satisfactory (mean 3.67, SD 0.75), with 57% (31/54) of the participants stating that their overall rating of the app was high or very high. Only 32% (17/54) of the participants reported having met their app use goals at the end of the study period.

# Discussion

## Principal Findings

There has been increasing international concern regarding the mental health and well-being of junior doctors [31]. Although there is emerging evidence that junior medical staff can be taught skills that will improve their mental health and reduce their risk of future mental health problems, the practical implementation of such training has proven to be a major barrier. This study represents the first ever evaluation of a smartphone app designed to meet the mental health and well-being needs of junior doctors. If shown to be effective, this type of app could be used widely at minimal cost and with minimal disruption. The results of this initial evaluation of the *Shift* app were mixed. There was evidence of a significant reduction in depression and anxiety symptoms. However, doctors' COVID-19 concerns regarding their personal safety significantly increased. Furthermore, there was no evidence of a significant reduction in work and social functioning scores, no decline in specific COVID-19 concerns for family and friends, and no increase in help seeking over the 30-day study period. In addition, although the app was positively rated by most junior doctors, only a quarter of the participants completed the poststudy assessment, making it difficult to deduce what the remainder of the participants thought of the app.

Further assessment examining a possible relation between app use variables and depressive and anxiety symptoms showed effects in the proposed direction; however, these analyses had low power and did not yield any significant results. Similarly, increases in COVID-19 pandemic safety concerns could not be explained by app use. This, together with the lack of a control group, makes it impossible to conclude that the reductions in depressive and anxiety symptoms or increases in COVID-19 pandemic concerns were because of engagement with the app. However, it needs to be noted that dosage considerations based on classical intervention studies, such as medication trials, do not always apply to studies using self-directed eHealth interventions because participants in this study could choose their own "dosage" targets as a part of the app onboarding procedure [32]. The self-prescriptive nature of *Shift* was based on extensive user consultations with junior doctors to facilitate use and acceptance of the app [22]. As such, participants were free to decide how often and for how long they would use the app over the study period. It is, therefore, possible that junior doctors only used the app to feel better at a specific point in time, which may have elicited transient effects that did not materialize at the postassessment measurement. In the engagement literature, the concept of e-attainment proposes that the users of digital mental health tools engage with the technology for as long as they need to in order to reach their desired mental health goals [33]. Doctors may have only used the app if they deemed the contents to be useful for improving their mental health or used the app in conjunction with other strategies outside of the digital arena. With regard to personal safety concerns because of the COVID-19 pandemic, it is possible that the increased hospitalization rates during the second COVID-19 pandemic wave in 2020, which coincided with the data collection period of this study, were a more important predictor of safety considerations than any potential alleviation of concerns that the app contents could provide.

Among those who fully participated in this study, including the poststudy assessment, the app was generally well received, with most users rating the app highly or very highly. However, because of the extensive dropout rate, these results may be skewed, given that study completion was associated with more favorable app ratings and higher app use. This, in turn, implies that the app did not substantially engage a large portion of the study participants. Upon the inspection of app use variables, it appears that in particular, the help-seeking, psychoeducational, and value-based components of the app were underutilized and therefore require further user consultation and revision. The present sample expressed willingness to seek help for psychological problems and one-fifth of the baseline sample indicated having sought help from a mental health professional in the last 4 weeks, which could explain why this section of the app was not of great interest. However, given that the promotion of early help seeking for mental health concerns is one of the key aims of the app, the failure to engage users to familiarize themselves with help-seeking options needs to be addressed before a wider dissemination of the app may be considered. To do this successfully, a variety of barriers to help seeking should be overcome, some of which relate to pervasive mental health stigma in the medical profession and fears of being reported to medical regulators, and others relate to situational factors such as time constraints and the pandemic context [9,17,34,35].

This study has several strengths and limitations that affect the interpretability of the current findings. The strengths of this research are that *Shift* is the first app of its kind to address work and other common stressors of junior doctors, a cohort that has been known to be subjected to intensive work demands and declining mental health trajectories during trainee years. The app was created in close collaboration with junior doctors and was tested extensively at each development step to ensure that the tool is appropriate, useful, and does not incur harm. In addition, in response to the COVID-19 pandemic, we anticipated heightened stress in this group. Therefore, the study design was modified to allow all eligible candidates to receive the app immediately for download after registering for the study. We also minimized the number of reminder messages to only 1 message to avoid adding to participants' stress by requesting their continued study participation. By avoiding these "push factors," the study became more naturalistic and consequently had greater ecological validity than originally planned.

Limitations relate to the study design and attrition, which decrease the validity of this research. Owing to the novel pandemic context, we were unable to adopt established measures of COVID-19 pandemic safety concerns for doctors when we designed the study. The self-generated COVID-19 items as well as our adjustments to the General Help-Seeking Questionnaire were not pretested before their inclusion in this investigation. As such, it is possible that using these measures limited the validity of this research. The removal of a control group effectively changed the robust design to an extensive pilot evaluation. Because the recruitment phase coincided with the COVID-19 pandemic, historical effects are likely confounders to the observations presented in this paper. It also needs to be noted that a small proportion of the study participants showed moderate or severe symptoms of depression or anxiety. It is, therefore, possible that generally better-adjusted doctors who would not perceive study participation as an undue burden on them agreed to partake in this investigation. As such, it is possible that doctors who would potentially benefit the most from exposure to mental health information did not seek out this study, which impacts the generalizability of findings and may also tap into a greater issue of hard-to-reach persons who already present with mental health problems. In addition, the high attrition rate of this research may have further introduced selection bias, and the app feedback questions may have been subject to participants' response bias. Although study completers did not vary in demographic characteristics and mental health profiles from those lost to follow-up, app acceptability and usability ratings, in particular, seem to have been skewed toward higher app use and more favorable app ratings among study completers. As for the high dropout rate, the analyses presented in this paper did not meet the initial expectations with regard to their achieved power. Therefore, it is possible that small but true effects may not have been detected. Future research should address these concerns and conduct a more robust examination of the effectiveness and overall usefulness of the app.

## Conclusions

Despite the crucial function that doctors play in ensuring the integrity and functioning of health care systems, doctors can feel undervalued and insecure in their ability to contribute what is needed to fulfill their role [36]. Considering the known barriers to addressing mental health concerns in this cohort, this study aimed to examine the usefulness of a tailored mental health app that can deliver initial mental health information and support options to junior doctors. Although the effectiveness of this approach is still in question, preliminary findings suggest that *Shift* seems to constitute a promising tool that may be able to reduce depressive and anxiety symptoms and improve outcomes for those who are willing to engage with digital mental health support. These results materialized despite a demanding pandemic context. Further research should seek to more robustly test the effectiveness for such apps on a wider scale among junior doctors. Although these types of individual interventions should constitute a part of the response to support physicians' mental health, they need to be a part of a broader suite of interventions that also consider the systemic and organizational factors that impact the health and well-being of health care workers.

## Acknowledgments

This research was funded by the New South Wales Ministry of Health and by the National Health and Medical Research Council's Investigator Grant awarded to the last author. The authors are grateful to Jennifer Chapman and Deborah Frew from the New South Wales Ministry of Health for their help in disseminating *Shift* study information through their channels and for their overall support of this research program. The authors also thank all junior doctors who participated in this research and value their contributions to the broader app development process. The development of *Shift* was made possible with funding from the New South Wales Ministry of Health and the iCare Foundation.

## Data Availability

Data from all participants were included in this paper, provided participants did not withdraw their informed consent to participate. We did not seek permission from participants to make deidentified data available as per the open science framework. Consequently, data are not publicly available because of ethical restrictions.

## Conflicts of Interest

Intellectual property for the *Shift* app is owned by the Black Dog Institute; however, the *Shift* app does not currently produce any income, and the authors do not receive any financial gain from this intellectual property.

Multimedia Appendix 1
Overview and categorization of *Shift* app activities.
[DOCX File , 18 KB - jmir_v24i9e38497_app1.docx ]

## References

1. Zhou AY, Panagioti M, Esmail A, Agius R, Van Tongeren M, Bower P. Factors associated with burnout and stress in trainee physicians: a systematic review and meta-analysis. JAMA Netw Open 2020 Aug 03;3(8):e2013761 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.13761] [Medline: 32809031]
2. Mata DA, Ramos MA, Bansal N, Khan R, Guille C, Di Angelantonio E, et al. Prevalence of depression and depressive symptoms among resident physicians: a systematic review and meta-analysis. JAMA 2015 Dec 08;314(22):2373-2383 [FREE Full text] [doi: 10.1001/jama.2015.15845] [Medline: 26647259]
3. Joyce C, Wang WC. Job satisfaction among Australian doctors: the use of latent class analysis. J Health Serv Res Policy 2015 Oct;20(4):224-230. [doi: 10.1177/1355819615591022] [Medline: 26079142]
4. Panagioti M, Geraghty K, Johnson J, Zhou A, Panagopoulou E, Chew-Graham C, et al. Association between physician burnout and patient safety, professionalism, and patient satisfaction: a systematic review and meta-analysis. JAMA Intern Med 2018 Oct 01;178(10):1317-1331 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3713] [Medline: 30193239]
5. Brooks SK, Chalder T, Gerada C. Doctors vulnerable to psychological distress and addictions: treatment from the Practitioner Health Programme. J Ment Health 2011 Apr 15;20(2):157-164. [doi: 10.3109/09638237.2011.556168] [Medline: 21542716]
6. Kumar S. Burnout and doctors: prevalence, prevention and intervention. Healthcare (Basel) 2016 Jun 30;4(3):37 [FREE Full text] [doi: 10.3390/healthcare4030037] [Medline: 27417625]
7. National mental health survey of doctors and medical students. Beyond Blue. URL: https://www.beyondblue.org.au/about-us/our-work-in-improving-workplace-mental-health/health-services-program/national-mental-health-survey-of-doctors-and-medical-students [accessed 2022-02-10]
8. Petrie K, Crawford J, Baker ST, Dean K, Robinson J, Veness BG, et al. Interventions to reduce symptoms of common mental disorders and suicidal ideation in physicians: a systematic review and meta-analysis. Lancet Psychiatry 2019 Mar;6(3):225-234. [doi: 10.1016/s2215-0366(18)30509-1]
9. Muhamad Ramzi NS, Deady M, Petrie K, Crawford J, Harvey S. Help-seeking for depression among Australian doctors. Intern Med J 2021 Dec 24;51(12):2069-2077. [doi: 10.1111/imj.15035] [Medline: 32833296]
10. Garelick AI. Doctors' health: stigma and the professional discomfort in seeking help. Psychiatrist 2018 Jan 02;36(3):81-84. [doi: 10.1192/pb.bp.111.037903]
11. Forbes MP, Iyengar S, Kay M. Barriers to the psychological well-being of Australian junior doctors: a qualitative analysis. BMJ Open 2019 Jun 12;9(6):e027558 [FREE Full text] [doi: 10.1136/bmjopen-2018-027558] [Medline: 31196900]
12. Reports and results. Medical Training Survey. 2021. URL: https://www.medicaltrainingsurvey.gov.au/Results/Reports-and-results [accessed 2022-02-10]
13. Whiteford HA, Harris MG, McKeon G, Baxter A, Pennell C, Barendregt JJ, et al. Estimating remission from untreated major depression: a systematic review and meta-analysis. Psychol Med 2013 Aug;43(8):1569-1585. [doi: 10.1017/S0033291712001717] [Medline: 22883473]

14. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. J Affect Disord 2014 Jan;152-154:45-51. [doi: 10.1016/j.jad.2013.10.002] [Medline: 24183486]
15. Kraus C, Kadriu B, Lanzenberger R, Zarate CA, Kasper S. Prognosis and improved outcomes in major depression: a review. Transl Psychiatry 2019 Apr 03;9(1):127 [FREE Full text] [doi: 10.1038/s41398-019-0460-3] [Medline: 30944309]
16. Clarke R, McKee M. Suicides among junior doctors in the NHS. BMJ 2017 May 26;357:j2527. [doi: 10.1136/bmj.j2527] [Medline: 28550153]
17. Bailey E, Robinson J, McGorry P. Depression and suicide among medical practitioners in Australia. Intern Med J 2018 Mar;48(3):254-258. [doi: 10.1111/imj.13717] [Medline: 29512324]
18. Worthington E, MacKenzie P. Doctor suicides prompt calls for overhaul of mandatory reporting laws. ABC News. 2017 Apr 13. URL: https://www.abc.net.au/news/2017-04-13/doctor-suicides-prompt-calls-for-overhaul/8443842 [accessed 2022-09-13]
19. Workforce Planning and Talent Development. JMO Wellbeing and Support Plan Implementation Report. Australia: NSW Ministry of Health; Nov 10, 2020.
20. Kushnir T, Malkinson R, Ribak J. Rational thinking and stress management in health workers: a psychoeducational program. Int J Stress Manag 1998;5(3):169-178. [doi: 10.1023/A:1022941031900]
21. Reeve A, Tickle A, Moghaddam N. Are acceptance and commitment therapy-based interventions effective for reducing burnout in direct-care staff? A systematic review and meta-analysis. Mental Health Rev J 2018 Jul 04;23(3):131-155. [doi: 10.1108/mhrj-11-2017-0052]
22. Counson I, Bartholomew A, Crawford J, Petrie K, Basarkod G, Moynihan V, et al. Development of the shift smartphone app to support the emotional well-being of junior physicians: design of a prototype and results of usability and acceptability testing. JMIR Form Res 2021 Dec 02;5(12):e26370 [FREE Full text] [doi: 10.2196/26370] [Medline: 34860662]
23. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
24. Guille C, Frank E, Zhao Z, Kalmbach DA, Nietert PJ, Mata DA, et al. Work-family conflict and the sex difference in depression among training physicians. JAMA Intern Med 2017 Dec 01;177(12):1766-1772 [FREE Full text] [doi: 10.1001/jamainternmed.2017.5138] [Medline: 29084311]
25. Spitzer RL, Kroenke K, Williams JB, 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]
26. Roy A, Druker S, Hoge EA, Brewer JA. Physician anxiety and burnout: symptom correlates and a prospective pilot study of app-delivered mindfulness training. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e15608 [FREE Full text] [doi: 10.2196/15608] [Medline: 32234708]
27. Khatun MF, Parvin MF, Rashid MM, Alam MS, Kamrunnahar M, Talukder A, et al. Mental health of physicians during COVID-19 outbreak in Bangladesh: a web-based cross-sectional survey. Front Public Health 2021 Feb 3;9:592058 [FREE Full text] [doi: 10.3389/fpubh.2021.592058] [Medline: 33634065]
28. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2002 May;180:461-464. [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]
29. Wilson C, Deane F, Ciarrochi J, Rickwoo D. Measuring help seeking intentions: properties of the General Help Seeking Questionnaire. Can J Counsel 2005;39(1):28 [FREE Full text]
30. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40.
31. Harvey SB, Epstein RM, Glozier N, Petrie K, Strudwick J, Gayed A, et al. Mental illness and suicide among physicians. Lancet 2021 Sep 04;398(10303):920-930. [doi: 10.1016/S0140-6736(21)01596-8] [Medline: 34481571]
32. Sanatkar S, Baldwin PA, Huckvale K, Clarke J, Christensen H, Harvey S, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019 Nov 28;21(11):e14728 [FREE Full text] [doi: 10.2196/14728] [Medline: 31778115]
33. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
34. Guille C, Speller H, Laff R, Epperson C, Sen S. Utilization and barriers to mental health services among depressed medical interns: a prospective multisite study. J Grad Med Educ 2010 Jun;2(2):210-214 [FREE Full text] [doi: 10.4300/JGME-D-09-00086.1] [Medline: 21975622]
35. She R, Wang X, Zhang Z, Li J, Xu J, You H, et al. Mental health help-seeking and associated factors among public health workers during the COVID-19 outbreak in china. Front Public Health 2021;9:622677 [FREE Full text] [doi: 10.3389/fpubh.2021.622677] [Medline: 34046387]
36. Riley GJ. Understanding the stresses and strains of being a doctor. Med J Aust 2004 Oct 04;181(7):350-353. [doi: 10.5694/j.1326-5377.2004.tb06322.x] [Medline: 15462646]

---

## Abbreviations

**CBT:** cognitive behavioral therapy

**EMM:** estimated marginal mean

---

*Edited by T Leung; submitted 05.04.22; peer-reviewed by P Matthews, A Abi-Jaoude; comments to author 16.08.22; revised version received 25.08.22; accepted 29.08.22; published 21.09.22.*

*Please cite as:*
*Sanatkar S, Counson I, Mackinnon A, Bartholomew A, Glozier N, Harvey S*
*Preliminary Investigation of Shift, a Novel Smartphone App to Support Junior Doctors' Mental Health and Well-being: Examination of Symptom Progression, Usability, and Acceptability After 1 Month of Use*
*J Med Internet Res 2022;24(9):e38497*
*URL: https://www.jmir.org/2022/9/e38497*
*doi:10.2196/38497*
*PMID:36129745*

Original Paper

# Long-term Weight Loss in a Primary Care–Anchored eHealth Lifestyle Coaching Program: Randomized Controlled Trial

Laura Hesseldal[1,2], BSc; Jeanette Reffstrup Christensen[1,3,4], PhD; Thomas Bastholm Olesen[5], PhD; Michael Hecht Olsen[2,6,7], PhD; Pernille Ravn Jakobsen[1], PhD; Ditte Hjorth Laursen[8], PhD; Jørgen Trankjær Lauridsen[9], PhD; Jesper Bo Nielsen[1], PhD; Jens Søndergaard[1], PhD; Carl Joakim Brandt[1], PhD

[1]Research Unit for General Practice, Department of Public Health, University of Southern Denmark, Odense, Denmark
[2]Steno Diabetes Center Zealand, Holdbaek, Denmark
[3]User Perspectives and Community-based Interventions, Department of Public Health, University of Southern Denmark, Odense, Denmark
[4]Research Unit of General Practice, Aarhus University, Aarhus, Denmark
[5]Steno Diabetes Center Odense, Odense University Hospital, Odense, Denmark
[6]Department of Internal Medicine, Holbaek Hospital, Holbaek, Denmark
[7]Department of Regional Health Research, University of Southern Denmark, Odense, Denmark
[8]Department of Public Health, University of Copenhagen, Copenhagen, Denmark
[9]Department of Economics and Data Science, University of Southern Denmark, Odense, Denmark

**Corresponding Author:**
Laura Hesseldal, BSc
Research Unit for General Practice
Department of Public Health
University of Southern Denmark
WP 9, J. B. Winsløws Vej 9A, 5000 Odense
Odense, 5000
Denmark
Phone: 45 314919945
Email: lhesseldal@health.sdu.dk

## Abstract

**Background:** Long-term weight loss in people living with obesity can reduce the risk and progression of noncommunicable diseases. Observational studies suggest that digital coaching can lead to long-term weight loss.

**Objective:** We investigated whether an eHealth lifestyle coaching program for people living with obesity with or without type 2 diabetes led to significant, long-term (12-month) weight loss compared to usual care.

**Methods:** In a randomized controlled trial that took place in 50 municipalities in Denmark, 340 people living with obesity with or without type 2 diabetes were enrolled from April 16, 2018, to April 1, 2019, and randomized via an automated computer algorithm to an intervention (n=200) or a control (n=140) group. Patients were recruited via their general practitioners, the Danish diabetes organization, and social media. The digital coaching intervention consisted of an initial 1-hour face-to-face motivational interview followed by digital coaching using behavioral change techniques enabled by individual live monitoring. The primary outcome was change in body weight from baseline to 12 months.

**Results:** Data were assessed for 200 participants, including 127 from the intervention group and 73 from the control group, who completed 12 months of follow-up. After 12 months, mean body weight and BMI were significantly reduced in both groups but significantly more so in the intervention group than the control group (–4.5 kg, 95% CI –5.6 to –3.4 vs –1.5 kg, 95% CI –2.7 to –0.2, respectively; $P<.001$; and –1.5 kg/m$^2$, 95% CI –1.9 to –1.2 vs –0.5 kg/m$^2$, 95% CI –0.9 to –0.1, respectively; $P<.001$). Hemoglobin $A_{1c}$ was significantly reduced in both the intervention (–6.0 mmol/mol, 95% CI –7.7 to –4.3) and control (–4.9 mmol/mol, 95% CI –7.4 to –2.4) groups, without a significant group difference (all $P>.46$).

**Conclusions:** Compared to usual care, digital lifestyle coaching can induce significant weight loss for people living with obesity, both with and without type 2 diabetes, after 12 months.

**Trial Registration:** ClinicalTrials.gov NCT03788915; https://clinicaltrials.gov/ct2/show/NCT03788915

*(J Med Internet Res 2022;24(9):e39741)* doi:10.2196/39741

**KEYWORDS**

obesity; digital behavioral coaching; health behavior change; interactive advice; lifestyle change; mobile intervention; patient engagement; telemedicine; eHealth; digital health; digital coach; weight loss; coaching; training; engagement; behavior changes; diabetes; type 2 diabetes; T2D

# *Introduction*

Long-term weight loss can reduce the risk, postpone the onset, and reduce the progression of noncommunicable diseases (NCDs) [1,2]. Numerous studies have shown that type 2 diabetes (T2D) can be slowed, halted, or even reversed through lifestyle changes, such as a low-calorie diet and increased physical activity [2,3]. This can lead to fewer long-term complications and probably a prolonged life expectancy [4]. Unfortunately, it is difficult for people living with obesity and T2D to achieve and maintain long-term weight loss [5]. Despite an intensive focus on T2D in general practice in Denmark, many patients are not treated optimally, nor do they follow recommendations for a healthy lifestyle [6]. Even though general practice is meant to support self-management and a healthy lifestyle, studies have shown that annual consultations seldom address lifestyle issues [7].

Meta-analyses and systematic reviews show that electronic health (eHealth) and mobile health (mHealth) solutions are significantly better than usual care, defined as routine diabetes self-care with no personalized feedback, at supporting weight loss in the short term (ie, within 3 to 6 months) for people living with obesity [8,9]. Behavior change techniques (BCTs) are an essential component of effective solutions. These involve automated, semi-automated, or human digital feedback [9,10]. Human feedback, particularly from health care professionals (HCPs), is most effective [11].

As described in detail in the study protocol [12], this study's collaborative eHealth tool, called LIVA, has been developed based on the experiences of approximately 140,000 individuals who used the collaborative eHealth tool (version 1.0) over a period of 15 years [12]. Version 1.0 has been further developed into version 2.0 based on feedback from patients, general practitioners, and HCPs [13-15]. HCPs use the eHealth tool to conduct digital lifestyle coaching as a 1-hour-long, physical or virtual, face-to-face motivational interview. The participant and the HCP collaborate and agree on goals for relevant lifestyle activities, such as diet and exercise, that the patient is motivated to improve [12].

However, there is limited evidence on the potential for such solutions to lead to weight loss over the long term (ie, longer than 12 months) [8,9]. In this randomized controlled trial (RCT), we aimed to investigate whether digital coaching through a multifaceted eHealth tool could help people living with obesity, with or without T2D, to achieve and sustain more significant long-term weight loss than an equivalent control group receiving usual care.

# *Methods*

## Study Design and Ethical Approval

This study was part of an RCT that took place in 2 of the 5 regions in Denmark: the Capital Region of Denmark, with 28 municipalities, and the Region of Southern Denmark, with 22 municipalities. The study was carried out from April 2019 to October 2021. The study was approved by the scientific committee of the Region of Southern Denmark (S-20170183G). All methods are described in detail in the study protocol [12]. The study is registered at ClinicalTrials.gov (NCT03788915).

## Participants

In Denmark, lifestyle support is managed by local municipalities at health care centers. For this study, participants in municipal lifestyle programs within the participating regions were recruited through their local health care centers, general practitioners (GPs), the Danish diabetes organization, and social media. Participants who expressed a desire to participate could then register at the eHealth tool website [12]. After registration, participants were contacted by telephone by a research assistant, who ensured that the participant met the inclusion criteria for BMI (30-45 kg/$m^2$) and age (18-70 years). The exclusion criteria were (1) a lack of internet access through a computer or smartphone, (2) pregnancy or planned pregnancy, and (3) presence of a serious or life-threatening disease, defined as a condition with less than a 1-year life expectancy.

## Randomization

Participants were randomized to the intervention group, who received usual care and the digital lifestyle coaching, or a control group, who received only the usual care preferred by the patient and their doctor. Randomization occurred after the participants had completed the medical examination via an automated computer algorithm in groups of 10 at a 6:4 ratio, where 60% of the recruited participants were randomized to the intervention group and the remaining 40% were assigned to the control group; this method was based on a pilot RCT [16] and is described in our protocol article [12]. Randomization was controlled to ensure that 50% of participants in both the intervention group and control group would be people living with obesity who had not previously been diagnosed with T2D, and to ensure that the other 50% of participants in both the intervention group and control group would be people living with obesity who had been diagnosed with T2D. Blinding the participants, the research assistant, and the health coach who provided the lifestyle coaching to all the participants who received the intervention was not possible after randomization. The research assistant and health coach had no role in analyzing or interpreting the data.

### Procedures

At the baseline meeting, the participants gave written informed consent and informed the research assistant about their use of medication. Afterwards, a brief medical examination was performed. The examination included measurements such as the participants' height, measured in centimeters, without shoes; weight, measured with clothes but without shoes (we subtracted 1 kg for clothing); waist and hip circumference, measured with a tape measure around the waist, between the lower rib and pelvic curvature and hip, with one hand above the inguinal medial line (in keeping with the European Health Examination Survey guideline [17]); and blood pressure, measured in a seated position after 10 minutes of rest without speaking, using an electronic, automatic blood pressure monitor (Omron Model M3). Three blood pressure measurements were performed 1 minute apart, and the lowest measured value was recorded [18]. Hemoglobin $A_{1c}$ ($HbA_{1c}$), total cholesterol, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglyceride (TG) levels were measured and assessed using finger-stick sampling with a device (Hemocue $HbA_{1c}$ 501 Analyzer) that can measure $HbA_{1c}$ in nonfasting blood samples [19]. To ensure the accuracy of the measurements, the Hemocue Analyzer was calibrated daily according to the manufacturer's instructions. Additional calibration was done monthly using a special kit to test the sensitivity and specificity of the measurements. A strict protocol was followed for the collection of blood samples. This examination was performed at both 6 and 12 months.

All participants filled out the European Quality of Life-5 Dimensions (EQ-5D-5L) (an online questionnaire on sociodemographic characteristics) and the Short-Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) [20,21]. The EQ-5D-5L descriptive system has 5 dimensions: mobility, self-care, usual activities, pain and discomfort, and anxiety and depression. Each dimension has 5 response levels: "no problems," "slight problems," " moderate problems," "severe problems," and "unable to/extreme problems." Responses are coded as single-digit numbers expressing the severity level selected for each dimension, which are then coded into a score ranging from 0.35 to 1.0. The SWEMWBS is a 7-item scale covering subjective well-being and psychological functioning. Each item is answered on a 5-point Likert scale, including "none of the time," "rarely," "some of the time," "often," and "all the time." The summary index ranges from 7 to 35. Higher scores indicate higher well-being and psychological functioning [20,21]. The participants answered both questionnaires at baseline and after 6 and 12 months of follow-up.

### Intervention

After an initial 45-to-60-minute consultation with the HCP, the intervention group received the individualized digital lifestyle coaching and used the eHealth tool to complete daily records and to send remarks directly to the HCP. Based on individual goal setting created using the SMART (specific, measurable, attainable, relevant, timely) model [22], the health coach provided weekly asynchronous digital coaching for each participant that included inspiring them, commending them on goal attainment, and seeking to help them stay motivated [13]. The subsequent asynchronous eHealth coaching sessions were carried out once a week for the first 6 months and then once a month for the last 6 months, as maintenance. The eHealth tool application is further described in the Template of the Intervention Description and Replication (TIDieR) (Multimedia Appendices 1 and 2).

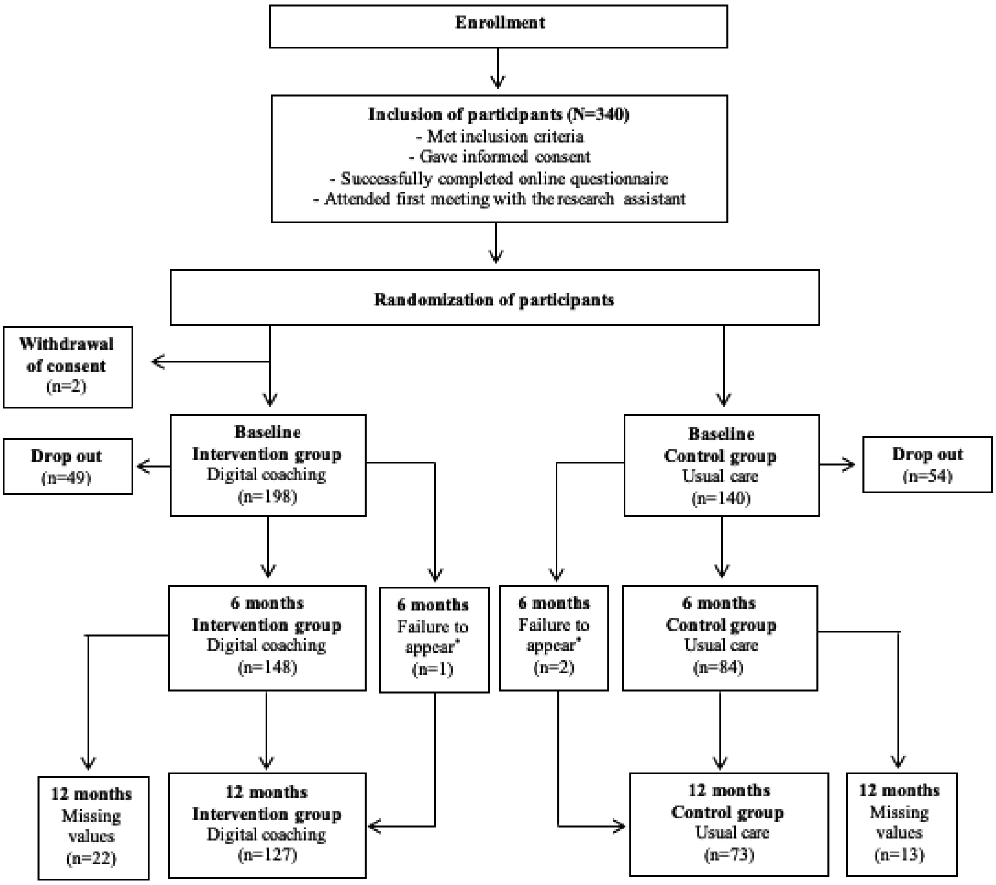
### Characteristics of HCPs

The HCPs who provided the digital lifestyle coaching through the eHealth tool were all educated as nurses, physiotherapists, dieticians, or occupational therapists. They all underwent special training in digital health coaching and had all practiced digital health coaching for at least 2 years. All participants were assigned a primary HCP so that there was a better chance of achieving a close and trusting professional relationship [12].

### Follow-up Procedure and COVID-19 Lockdown

After 6 and 12 months, the participants were invited to a brief medical examination, similar to the baseline examination, performed by a research assistant. To confirm patient-reported data, the same medical data were also retrieved from the shared medication record (abbreviated as "FMK" in Danish) and from laboratory results, measured at GP clinics. Participants were also asked to complete the same web-based questionnaire [12]. Participants were contacted 1 month before their 6- and 12-month assessments by telephone to schedule the assessment. If a participant did not respond, a voice mail was left explaining the purpose of the call. Another telephone call was made a week later and again 1 month later. Participants who had not responded to 4 different attempts were considered lost to follow-up. Due to the COVID-19 lockdown and national restrictions, some participants could not attend their 12-month assessment. Therefore, the 12-month assessment period was extended by 4 months, so that follow-up after baseline also covered 12 to 16 months. However, this extension might not have been sufficient to obtain 12-month follow-up data from all participants (Figure 1). Thus, this paper reports 12-month follow-up data from 126 participants in the intervention group and 71 participants in the control group who attended follow-up examinations at both 6 and 12 months.

**Figure 1.** Flowchart of participation in 12-month randomized controlled trial. *Failed to appear at the assessment but appeared at the next follow-up.



### Outcomes

The primary outcome was reduction in mean body weight (BW), assessed as the difference between BW at baseline and at 12 months, and as the difference divided by baseline BW. The proportion of participants who had significant weight loss (ie, >5% of baseline BW [23]) was also assessed at 6 and 12 months. Our secondary outcome was the change in $HbA_{1c}$ at 6 and 12 months compared to baseline. The tertiary outcomes were body composition (BMI and waist/hip ratio), lipid levels (total cholesterol, LDL-C, HDL-C, and TG), blood pressure (systolic and diastolic), and changes in mental health and quality of life. All differences in tertiary outcomes were calculated from baseline to 6 months and baseline to 12 months.

### Statistical Analysis

All analyses used Stata/BE (version 17.0; StataCorp) and were performed on data from participants who attended the 12-month follow-up; other participants were considered dropouts. Baseline characteristics of all participants allocated to the intervention and control groups were analyzed descriptively. The statistical significance of differences in baseline characteristics of the participants who attended the 12-month follow-up was assessed with the Student *t* test and the Kruskal-Wallis test. The statistical significance of between-group differences in outcomes at 6 and 12 months was assessed with either a 1-way ANOVA or the chi-square test. Statistical significance was set at 2-tailed $P<.05$. In addition, we performed a per protocol analysis by using only data from participants who had been using the eHealth tool for 365 days or more. Finally, we performed a regression analysis that included an interaction term to determine whether participants with T2D responded differently to the intervention.

## Results

### Participant Characteristics

From April 16, 2018, to April 1, 2019, 340 participants were randomized. Two participants in the intervention group decided to withdraw their consent; thus, a total of 338 participants were included. The intervention group included 198 participants (128/198 female, 65%) and the control group included 140 participants (85/140 female, 61%) (Figure 1). At baseline, the intervention and control groups were comparable (Table 1). Participants' mean BW was 103.7 kg, their mean BMI was 35.3 $kg/m^2$, and their mean $HbA_{1c}$ was 6.6% (Table 1).

A total of 200 participants completed the 12-month follow-up (Figure 1). Participants who dropped out of the study (ie, did not complete the 12-month follow-up) were generally not different from the active participants. However, significantly fewer participants who dropped out were married, more were unmarried (including divorce), and they had slightly higher diastolic blood pressure and lower quality of life (Multimedia Appendix 3), although there were no significant differences within the intervention group or the control group. At baseline, there were significantly fewer participants receiving metformin, SGLT2, or calcium antagonists in the group of participants who dropped out (Multimedia Appendix 3).

**Table 1.** Baseline characteristics of participants.

| Characteristics | Intervention group (n=127) | Control group (n=73) | Total (N=200) |
|---|---|---|---|
| Age (years), mean (SD) | 52.3 (10) | 52.3 (12) | 52.3 (11) |
| **Sex, n (%)** | | | |
| Female | 86 (68) | 41 (56) | 127 (64) |
| Male | 41 (32) | 32 (44) | 73 (37) |
| **Diabetes, n (%)** | | | |
| Yes | 62 (50) | 36 (49) | 98 (49) |
| No | 65 (51) | 37 (51) | 102 (51) |
| **Education, n (%)** | | | |
| None | 19 (15) | 15 (21) | 34 (17) |
| Short | 33 (26) | 19 (26) | 52 (26) |
| Medium | 61 (48) | 30 (41) | 91 (46) |
| Long | 12 (9) | 9 (12) | 21 (11) |
| Don’t know | 2 (2) | 0 (0) | 2 (1) |
| **Marital status, n (%)** | | | |
| Married | 92 (72) | 49 (67) | 141 (71) |
| Unmarried[a] | 33 (26) | 23 (32) | 56 (28) |
| Widowed | 2 (2) | 1 (1) | 3 (2) |
| **Occupational status, n (%)** | | | |
| Employed | 96 (76) | 48 (66) | 144 (72) |
| Out of work[b] | 10 (8) | 6 (8) | 16 (8) |
| Retired | 20 (16) | 17 (23) | 37 (19) |
| Student | 1 (0) | 2 (3) | 3 (2) |
| **Body composition, mean (SD)** | | | |
| Weight (kg) | 103.0 (15.7) | 104.9 (15.8) | 103.7 (15.7) |
| BMI (kg/m$^2$) | 34.8 (3.7) | 36.0 (3.8) | 35.3 (3.8) |
| Hip circumference (cm) | 121.1 (9.6) | 121.7 (10.2) | 121.3 (9.8) |
| Waist circumference (cm) | 117.7 (11.4) | 121.2 (11.7) | 119.0 (11.6) |
| Waist to hip ratio | 1.0 (0.1) | 1.0 (0.1) | 1.0 (0.1) |
| **Glycemic control** | | | |
| $HbA_{1c}$[c] (%), mean (SD) | 6.6 (1.3) | 6.6 (1.3) | 6.6 (1.3) |
| $HbA_{1c}$ (mmol), mean (SD) | 48.3 (13.6) | 48.4 (14.0) | 48.3 (13.7) |
| $HbA_{1c}$ <6.5%, n (%) | 70 (55) | 41 (56) | 111 (56) |
| **Blood pressure, mean (SD)** | | | |
| Systolic (mm Hg) | 130.6 (13.8) | 131.4 (16.6) | 130.9 (14.8) |
| Diastolic (mm Hg) | 86.0 (8.1) | 86.5 (10.4) | 86.2 (9.0) |
| **Lipids** | | | |
| Total cholesterol (mmol/l), mean (SD) | 4.9 (1.3) | 4.8 (1.1) | 4.9 (1.2) |
| High density lipoprotein (mmol/l), median (IQR) | 1.2 (0.7) | 1.2 (0.6) | 1.2 (0.5) |
| Low density lipoprotein (mmol/l), median (IQR) | 2.3 (1.4) | 2.2 (1.6) | 2.2 (1.5) |
| Triglycerides (mmol/l), median (IQR) | 2.6 (2.2) | 2.7 (2.5) | 2.6 (2.3) |
| Mental Health score[d] , mean (SD) | 24.9 (3.2) | 24.5 (3.9) | 24.8 (3.5) |

| Characteristics | Intervention group (n=127) | Control group (n=73) | Total (N=200) |
|---|---|---|---|
| Quality of life score[e], mean (SD) | 0.8 (0.1) | 0.8 (0.1) | 0.8 (0.1) |

[a] Single or divorced.

[b] On maternity leave or receiving unemployment or cash benefits.

[c] $HbA_{1c}$: hemoglobin $A_{1c}$.

[d] Measured with Short-Warwick-Edinburgh Mental Well-being Scale; index ranges from 7-35.

[e] Index calculated based on the EQ-5D-5L; ranges from 0.35 to 1.0.

## Primary Outcome

At the 6-month follow-up, BW was significantly reduced in the intervention group (–4.5 kg, 95% CI –5.4 to –3.5) and not significantly reduced in the control group (–0.3 kg, 95% CI –1.1 to 0.4). This between-group difference was statistically significant (*P*<.001) (Table 2). Our primary outcome, BW at the 12-month follow-up, was significantly reduced in both the intervention group (–4.5 kg, 95% CI –5.6 to –3.4) and the control group (–1.5 kg, 95% CI –2.7 to –0.2); the reduction in the intervention group was significantly greater (*P*<.001) (Table 3). There was a significant weight loss (defined as >5% BW, *P*=.01) in a greater proportion of participants in the intervention group (48/127, 37.8%) than the control group (14/73,19%). The same pattern was seen among the per protocol participants (ie, the participants who used the eHealth tool for 365 days or more).

Within the intervention group, the effect over time on BW reduction was equal in participants with and without T2D, but in the control group, participants without T2D did not achieve significant weight change (Figure 2). Between the 6- and 12-month follow-ups, there was a significant weight reduction in the control group participants with T2D. All other weight changes at the 6- and 12-month follow-ups were not significant (Figure 2).

**Table 2.** Between-group differences in changes from baseline to the 6-month follow-up. Results in italics represent a significant change from baseline.

| Characteristics | Intervention group at 6 months (n=126) | Control group at 6 months (n=71) | Between-group difference (95% CI) | *P* value |
|---|---|---|---|---|
| **Weight** | | | | |
| Change vs baseline (kg), mean (95% CI) | *–4.5* (–5.4 to –3.5) | –0.3 (–1.1 to 0.4) | 4.2 (2.8 to 5.5) | <.001 |
| Change vs baseline (%), mean (95% CI) | *–4.4* (–5.3 to –3.4) | –0.4 (–1.1 to 0.3) | 3.9 (2.6 to 5.3) | <.001 |
| Lost >5% bodyweight (n), % | 49 (38.9) | 6 (8.5) | 30.4 (19.7 to 41.1) | <.001 |
| **Hemoglobin $A_{1c}$** | | | | |
| Change (%), mean (95% CI) | *–0.5* (–0.6 to –0.3) | *–0.4* (–0.5 to –0.2) | 0.1 (–0.2 to 0.4) | .49 |
| Change (mmol/mol), mean (95% CI) | *–4.8* (–6.7 to –3.0) | *–3.8* (–5.9 to –1.8) | 1.0 (–1.8 to 3.9) | .49 |
| Reduction from >6.5% to <6.5% (only in T2D patients), n/N (%)[a] | 22/63 (35) | 9/34 (27) | 8.4 (–10.4 to 27.3) | .39 |
| **Body composition** | | | | |
| BMI change (kg/m$^2$), mean (95% CI) | *–1.5* (–1.8 to –1.2) | –0.1 (–0.4 to 0.1) | 1.4 (0.9 to 1.8) | <.001 |
| Change in hip circumference (cm), mean (95% CI) | *–5.5* (–6.5 to –4.6) | *–1.9* (–3.1 to –0.7) | 3.6 (2.0 to 5.2) | <.001 |
| Change in waist circumference (cm), mean (95% CI) | *–8.9* (–10.2 to –7.7) | *–3.3* (–4.8 to –1.8) | 5.6 (3.6 to 7.6) | <.001 |
| Change in waist/hip ratio (cm), mean (95% CI) | *–0.030* (–0.041 to –0.019) | –0.012 (–0.026 to 0.002) | 0.018 (–0.000 to 0.036) | .052 |
| **Blood pressure** | | | | |
| Change in systolic pressure (mm Hg), mean (95% CI) | –1.4 (–3.6 to 0.8) | –0.3 (–3.4 to 2.9) | 1.1 (–2.6 to 4.9) | .56 |
| Change in diastolic pressure (mm Hg), mean (95% CI) | *–2.0* (–3.2 to –0.7) | –0.8 (–2.5 to 1.0) | 1.2 (–0.9 to 3.3) | .27 |
| **Lipids** | | | | |
| Change in total cholesterol (mmol/ml), mean (95% CI) | –0.2 (–0.3 to 0.0) | 0.1 (–0.1 to 0.3) | 0.3 (–0.0 to 0.5) | .07 |
| Change in high density lipoprotein (mmol/ml), median (95% CI) | *–0.1* (–0.2 to –0.0) | –0.1 (–0.1 to 0.0) | 0.0 (–0.0 to 0.1) | .51 |
| Change in triglyceride (mmol/ml), median (95% CI) | *–0.6* (–0.9 to –0.3) | 0.7 (–1.6 to 3.1) | 1.3 (–0.6 to 3.1) | .17 |
| Change in low density lipoprotein (mmol/ml), median (95% CI)[a] | 0.2 (–0.0 to 0.4) | *0.4* (0.1 to 0.6) | 0.2 (–0.1 to 0.5) | .22 |
| Change in quality of life score, mean (95% CI) | 0.0 (–0.0 to 0.0) | –0.0 (–0.0 to 0.0) | –0.0 (–0.0 to 0.0) | .14 |
| Change in mental health score, mean (95% CI) | –0.3 (–0.9 to 0.3) | 0.3 (–0.6 to 1.2) | 0.6 (–0.5 to 1.6) | .27 |

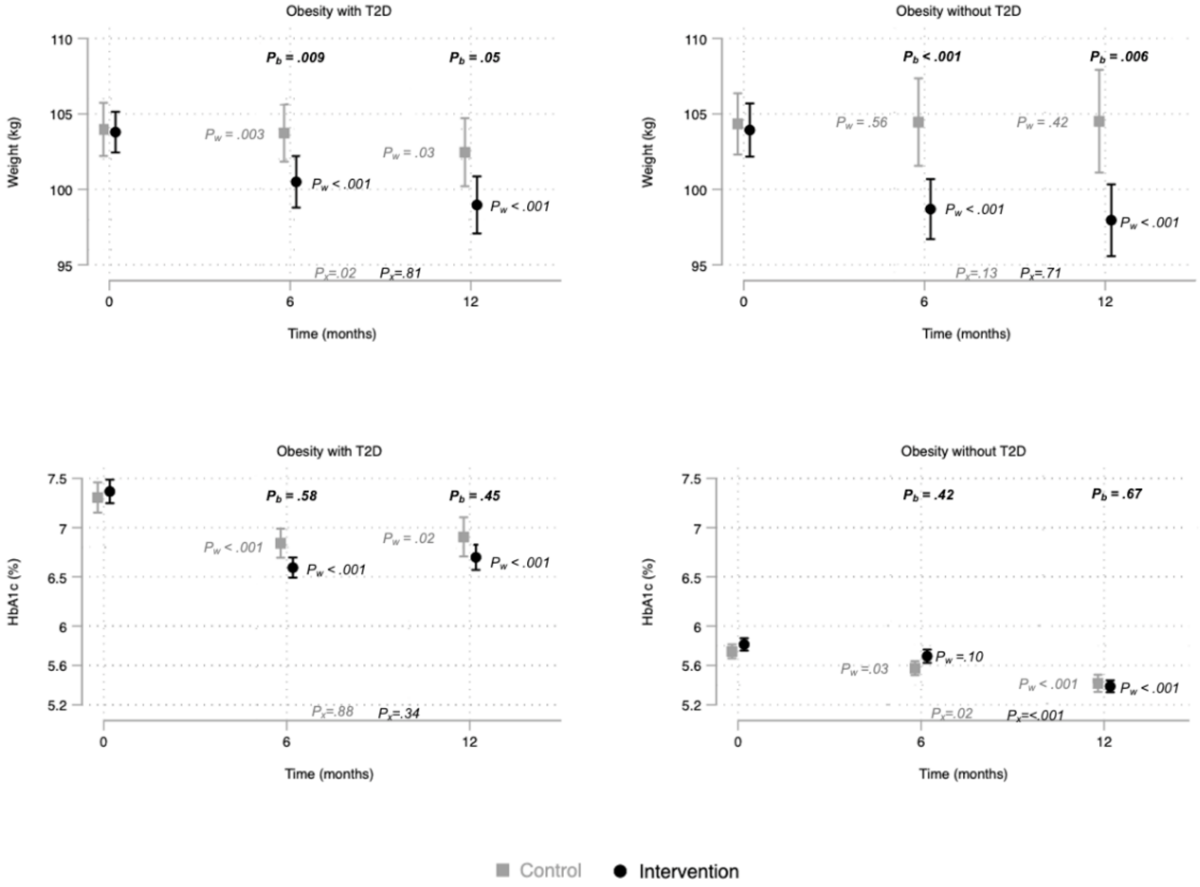
[a] Calculated in 153/200 participants, including 95/127 in the intervention group and 59/73 in the control group.

**Table 3.** Between-group differences in changes from baseline to the 12-month follow-up. Results in italics represent a significant change from baseline.

| Characteristics | Intervention group at 12 months (n=127) | Control group at 12 months (n=73) | Between-group difference, (95% CI) | *P* value |
|---|---|---|---|---|
| **Weight** | | | | |
| Change vs baseline (kg), mean (95% CI) | *–4.5* (–5.6 to –3.4) | *–1.5* (–2.7 to –0.2) | 3.0 (1.3 to 4.8) | <.001 |
| Change vs baseline (%), mean (95% CI) | –4.6 (–5.7 to –3.4) | *–1.4* (–2.6 to –0.1) | 3.2 (1.4 to 5.0) | <.001 |
| Lost >5% bodyweight (n), % | 37.8 (48) | 19.2 (14) | 18.6 (6.2 to 30.9) | .01 |
| **Hemoglobin $A_{1c}$** | | | | |
| Change (%), mean (95% CI) | *–0.5* (–0.7 to –0.4) | *–0.4* (–0.7 to –0.2) | 0.1 (–0.2 to 0.4) | .41 |
| Change (mmol/mol), mean (95% CI) | *–6.0* (–7.7 to –4.3) | *–4.9* (–7.4 to –2.4) | 1.0 (–1.9 to 4.0) | .46 |
| Reduction from >6.5% to <6.5% (only in in T2D patients), n/N (%)[a] | 22/62 (36) | 10/36 (28) | 7.7 (–11.1 to 26.5) | .43 |
| **Body composition** | | | | |
| BMI change ($kg/m^2$), mean (95% CI) | *–1.5* (–1.9 to –1.2) | *–0.5* (–0.9 to –0.1) | 1.0 (0.4 to 1.7) | <.001 |
| Change in hip circumference (cm), mean (95% CI) | *–5.9* (–7.0 to –4.8) | *–2.4* (–3.8 to –1.0) | 3.5 (1.7 to 5.3) | <.001 |
| Change in waist circumference (cm), mean (95% CI) | *–9.9* (–11.3 to –8.4) | *–4.5* (–6.6 to –2.5) | 5.3 (2.8 to 7.8) | <.001 |
| Change in waist/hip ratio (cm), mean (95% CI) | *–0.036* (–0.047 to –0.024) | *–0.019* (–0.036 to –0.002) | 0.016 (0.003 to 0.0361) | .11 |
| **Blood pressure** | | | | |
| Change in systolic pressure (mm Hg), mean (95% CI) | *–3.3* (–5.3 to –1.4) | *–4.7* (–8.0 to –1.3) | –1.3 (–5.0 to 2.3) | .47 |
| Change in diastolic pressure (mm Hg), mean (95% CI) | *–2.4* (–3.6 to –1.2) | –1.4 (–3.7 to 0.9) | 1.0 (–1.4 to 3.4) | .40 |
| **Lipids** | | | | |
| Change in total cholesterol (mmol/ml), mean (95% CI) | *–0.4* (–0.5 to –0.2) | *–0.2* (–0.5 to –0.0) | 0.1 (–0.2 to 0.4) | .42 |
| Change in high density lipoprotein, (mmol/ml), median, mean (95% CI) | 0.6 (–1.0 to 2.2) | *–0.2* (–0.3 to –0.2) | –0.8 (–2.9 to 1.3) | .44 |
| Change in triglycerides (mmol/ml), median, (95% CI) | *–0.8* (–1.1 to –0.6) | *–0.8* (–1.1 to –0.5) | 0.05 (–0.4 to 0.5) | .81 |
| Change in low density lipoprotein (mmol/ml), median, (95% CI)[a] | 0.2 (0.0 to 0.3) | 0.3 (0.0 to 0.5) | 0.1 (–0.2 to 0.4) | .58 |
| Change in quality of life score, mean (95% CI) | 0.0 (–0.0 to 0.0) | –0.0 (–0.0 to 0.0) | –0.0 (–0.0 to 0.0) | .47 |
| Change in mental health score, mean (95% CI) | 0.4 (–0.2 to 1.0) | 0.3 (–0.6 to 1.2) | –0.1 (–1.1 to 0.9) | .84 |

[a] Calculated in 153/200 participants, including 95/127 in the intervention group and 59/73 in the control group.

**Figure 2.** Body weight and hemoglobin $A_{1c}$ at baseline, at the 6-month follow-up (n=197), and at the 12-month follow-up (n=200) in subgroups with and without type 2 diabetes. Dots indicate the mean and lines indicate the standard error of the mean. $P_w$: *P* value for changes from baseline within groups; $P_b$: *P* value for changes from baseline between groups; $P_x$: *P* value for changes from 6 to 12 months within groups; $HbA_{1c}$: hemoglobin $A_{1c}$; T2D: type 2 diabetes.



## Secondary Outcome

At the 6-month follow-up, our secondary outcome, $HbA_{1c}$, was equally reduced in both groups. At the 12-month follow-up, $HbA_{1c}$ in both the intervention group (–0.5%, 95% CI –0.7 to –0.4) and the control group (–0.4%, 95% CI –0.7 to –0.2) were still equally reduced (Table 3). The largest reduction of $HbA_{1c}$ in the intervention group was seen within the first 6 months (Figure 2). Although Figure 2 might seem to indicate that the intervention only reduced $HbA_{1c}$ in participants with T2D, T2D did not interact with the effect of the intervention on $HbA_{1c}$ (all values: *P*>.43). The reduction in $HbA_{1c}$ at the 12-month follow-up among participants in the intervention group with T2D was greater (–0.7%, 95% CI –1.1 to –0.4) than the reduction in participants without T2D (–0.4%, 95% CI –0.4 to –0.3). From the 6-month follow-up to the 12-month follow-up, participants without T2D significantly reduced $HbA_{1c}$, but there were no significant changes within the group of participants with T2D (Figure 2). The proportion of participants in the intervention group whose $HbA_{1c}$ became normal was significantly greater at the 12-month follow-up (54/27, 43%) than at baseline (33/127, 25.9%), but $HbA_{1c}$ becoming normal was not significantly more prevalent than in the control group (Multimedia Appendix 3).

## Tertiary Outcomes

At the 6- and 12-month follow-ups, mean BMI decreased significantly in both groups, but significantly more so in the intervention group (–1.5 kg/m$^2$, 95% CI –1.9 to –1.2 vs –0.5 kg/m$^2$, 95% CI –0.9 to –0.1; *P*<.001). The waist/hip ratio was reduced significantly in both groups, but there was no significant between-group difference (Table 3).

At the 12-month follow-up, blood pressure, total cholesterol, and TG were reduced in both groups without any between-group differences (Table 3). HDL-C was decreased in both groups at the 6-month follow-up. At the 12-month follow-up, HDL-C was still decreased in the control group but was nonsignificantly increased in the intervention group. There were no statistically significant between-group differences at either the 6- or 12-month follow-ups.

At both the 6- and 12-month follow-ups, quality of life and mental health were unchanged in both groups (Tables 2 and 3). In general, medications (assessed as the defined daily dose for glucose-lowering and blood pressure–lowering drugs) did not change in any of the groups. However, use of dipeptidyl peptidase-4 inhibitors (DPP4s) decreased significantly in the intervention group, while use increased in the control group (*P*=.03) (Multimedia Appendix 3). Use of angiotensin-converting enzyme inhibitors increased insignificantly in the intervention group and decreased, although not significantly, in the control group (*P*=.06) (Multimedia Appendix 3).

# Discussion

## Principal Results

The main objective of this 12-month RCT was to see if individualized digital lifestyle coaching, enabled by an eHealth tool, could help people living with obesity with or without T2D to achieve and maintain a significant weight loss. This objective was met, with a mean weight loss of 4.5 kg in the intervention group, compared to 1.5 kg in the control group, after 12 months of follow-up.

## Comparison With Prior Work

These findings support previous studies that used eHealth solutions to promote lifestyle changes [24-26]. Our results are in line with a recent meta-analysis of studies using in-person behavioral counseling together with an eHealth intervention that showed a BW reduction of –4.65% [25].

The beneficial effect of the intervention in our study was probably due to the combination of face-to-face coaching and asynchronous eHealth with a different BCT, which has been proven effective in other studies [27,28]. The initial establishment of an honest and trustworthy relationship was found to be relevant in the qualitative interviews that the research team conducted while developing this study’s eHealth tool [13-15]. This finding is supported by smaller studies demonstrating that patients who found lifestyle changes challenging appeared to improve health behaviors when they used digital coaching that built on an empathetic relationship [16]. Observational retrospective studies suggest digital eHealth intervention incorporating personal coaching and BCTs may promote weight loss better over a 12-month period compared to studies with either face-to-face coaching or eHealth alone [29,30]. However, RCTs of eHealth solutions providing individualized coaching with follow-up at 12 months are sparse [31,32].

From baseline to the 12-month follow-up, $HbA_{1c}$ was reduced in both the intervention and control groups without a significant difference between the groups. This lack of difference was in contrast to a prior meta-analysis [26] and could not be explained by the small decrease in DPP4 use in the intervention group. The fact that the intervention did not reduce $HbA_{1c}$ significantly more in participants with T2D may reflect blood glucose already being well regulated at baseline in most of the participants with T2D (mean $HbA_{1c}$ was 6.6%). It is important that $HbA_{1c}$ was significantly reduced between the 6- and 12-month follow-ups in the participants without T2D in both groups, although BW was not reduced in the control group without T2D, suggesting that lifestyle changes other than weight reduction, such as more exercise, may have contributed to the $HbA_{1c}$ reduction.

Systolic blood pressure was significantly reduced in both groups at the 12-month follow-up, without a significant between-group difference, which might reflect the blood pressure reduction being a consequence of participation in the study (ie, the “healthy participator” effect) and only partly secondary to the weight loss. The same explanation is likely for the lipid findings. The lack of change in quality of life and mental health in both groups probably reflects these questionnaires being rather broad and therefore very robust toward changes in selective interventions. It is not unlikely that a specific overweight questionnaire would have picked up improvements related to the observed weight loss, but there would probably not have been a significant difference between the groups.

## Limitations

The dropout rate at the 12-month follow-up was 138 of 338 (40.8%), which is similar to attrition rates reported in other studies [9]. Although this could have created attrition bias, the participants who dropped out were closely comparable to the participants who came to the 12-month follow-up, and the characteristics of the participants in the intervention and control groups who came to the 12-month follow-up did not differ from each other at baseline. This may reflect many of the dropouts occurring at random due to COVID-19 restrictions. However, the relatively high number of dropouts reduced the power, making a subgroup analysis of participants with T2D difficult, with a high risk of false negative results, and may explain the nonsignificant effect of the intervention on $HbA_{1c}$. Another limitation of this study was the number of participants who came to the 6-month follow-up but missed the 12-month follow-up, possibly due to COVID-19 restrictions. As our clinical end points needed physical attendance, it was not possible to follow up with participants who dropped out. However, we repeated the analysis with imputation used to replace missing values. For 3 participants, the 12-month data was used to impute missing 6-month data, and for 35 participants, the 6-month data was used to impute missing 12-month data. This analysis obtained similar results. All the participants randomized to the intervention group who stayed in the study used the eHealth tool, indicating that if the eHealth tool is used in the future, the therapist will quickly be able to identify who is not satisfied with the individualized digital coaching. For these subjects, it will be possible for the therapist to recommend other treatment strategies.

Although evidence suggests that human feedback and coaching is an important element for success, our study design did not allow us to comment on the relative effectiveness of the components of this study’s eHealth approach. On the other hand, the randomized design of our study is a strength, showing that together with the relatively low cost of the intervention, a large scale-up seems possible.

## Conclusion

It is possible to induce and maintain lifestyle changes leading to significant and sustainable 12-month, long-term weight loss among people living with obesity with or without T2D using individualized digital lifestyle coaching, in comparison to usual care. These findings suggest that coaching with an eHealth tool based on real-time monitoring incorporating personal coaching and BCTs through smartphones can lead to improved lifestyles that may have the potential to further reduce the incidence and severity of NCDs.

## Acknowledgments

We thank Camilla Sortsø for her input on the protocol and the design of the study. We also thank all the participants for their willingness to take part in the study. None of the participants were financially compensated for their participation.

## Authors' Contributions

CJB, JBN, and JS were involved in designing the trial. CJB, JRC, JBN, DHL, and JS were responsible for implementing the trial. LH and TBO performed the statistical analysis. LH and JRC wrote the first draft of the report with input from TBO, MHO, and CJB. Afterwards, PRJ, DHL, JTL, JBN, and JS contributed input. All authors were involved in interpretation of the results. All authors contributed to, read, and approved the final manuscript.

## Conflicts of Interest

This study acquired no external funding. All authors were financially supported by their employer, the University of Southern Denmark, except DHL, who was financially supported by LIVA Healthcare A/S, which also paid for the coaching and instruments used in the study. A formal research agreement has been made between the University of Southern Denmark and LIVA Healthcare A/S to guarantee that LIVA Healthcare A/S cannot influence any results of the study, and in case there is any doubt, JS and the University of Southern Denmark have the final say. CJB is an original cofounder of LIVA Healthcare A/S and owns stock in LIVA Healthcare A/S, the company that developed parts of the technical platform. CJB works today at the Research Unit for General Practice at the University of Southern Denmark and as a consultant to LIVA Healthcare A/S. DHL is employed at LIVA Healthcare A/S. DHL primarily contributed information about the use of the LIVA app in clinical practice. LH, JRC, TBO, MHO, PRJ, JTL, JBN, and JS have no financial interests in LIVA Healthcare A/S or any other aspects of this study.

Multimedia Appendix 1
Template of the Intervention Description and Replication checklist for the LIVA 2.0 program.
[DOCX File , 17 KB - jmir_v24i9e39741_app1.docx ]

Multimedia Appendix 2
YouTube video showing an example of the Liva app used in real life.
[DOCX File , 13 KB - jmir_v24i9e39741_app2.docx ]

Multimedia Appendix 3
Supplementary tables and figures.
[DOCX File , 43 KB - jmir_v24i9e39741_app3.docx ]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1285 KB - jmir_v24i9e39741_app4.pdf ]

## References

1. Noncommunicable Diseases. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2022-08-21]
2. Lean ME, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet 2018 Feb 10;391(10120):541-551. [doi: 10.1016/S0140-6736(17)33102-1] [Medline: 29221645]
3. Johansen MY, MacDonald CS, Hansen KB, Karstoft K, Christensen R, Pedersen M, et al. Effect of an intensive lifestyle intervention on glycemic control in patients with type 2 diabetes: a randomized clinical trial. JAMA 2017 Aug 15;318(7):637-646 [FREE Full text] [doi: 10.1001/jama.2017.10169] [Medline: 28810024]
4. Gong Q, Zhang P, Wang J, Ma J, An Y, Chen Y, Da Qing Diabetes Prevention Study Group. Morbidity and mortality after lifestyle intervention for people with impaired glucose tolerance: 30-year results of the Da Qing Diabetes Prevention Outcome Study. Lancet Diabetes Endocrinol 2019 Jun;7(6):452-461 [FREE Full text] [doi: 10.1016/S2213-8587(19)30093-2] [Medline: 31036503]
5. Hall KD, Kahan S. Maintenance of lost weight and long-term management of obesity. Med Clin North Am 2018 Jan;102(1):183-197 [FREE Full text] [doi: 10.1016/j.mcna.2017.08.012] [Medline: 29156185]
6. Bo A, Thomsen RW, Nielsen JS, Nicolaisen SK, Beck-Nielsen H, Rungby J, et al. Early-onset type 2 diabetes: Age gradient in clinical and behavioural risk factors in 5115 persons with newly diagnosed type 2 diabetes-Results from the DD2 study. Diabetes Metab Res Rev 2018 Mar;34(3):10.1002/dmrr.2968. [doi: 10.1002/dmrr.2968] [Medline: 29172021]

7. du Pon E, Wildeboer AT, van Dooren AA, Bilo HJG, Kleefstra N, van Dulmen S. Active participation of patients with type 2 diabetes in consultations with their primary care practice nurses - what helps and what hinders: a qualitative study. BMC Health Serv Res 2019 Nov 08;19(1):814 [FREE Full text] [doi: 10.1186/s12913-019-4572-5] [Medline: 31703680]
8. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
9. Sherrington A, Newham JJ, Bell R, Adamson A, McColl E, Araujo-Soares V. Systematic review and meta-analysis of internet-delivered interventions providing personalized feedback for weight loss in overweight and obese adults. Obes Rev 2016 Jun;17(6):541-551 [FREE Full text] [doi: 10.1111/obr.12396] [Medline: 26948257]
10. Gong E, Baptista S, Russell A, Scuffham P, Riddell M, Speight J, et al. My Diabetes Coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: randomized effectiveness-implementation trial. J Med Internet Res 2020 Nov 05;22(11):e20322 [FREE Full text] [doi: 10.2196/20322] [Medline: 33151154]
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 Aug;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
12. Brandt CJ, Christensen JR, Lauridsen JT, Nielsen JB, Søndergaard J, Sortsø C. Evaluation of the clinical and economic effects of a primary care anchored, collaborative, electronic health lifestyle coaching program in denmark: protocol for a two-year randomized controlled trial. JMIR Res Protoc 2020 Jun 25;9(6):e19172 [FREE Full text] [doi: 10.2196/19172] [Medline: 32584260]
13. Brandt CJ, Clemensen J, Nielsen JB, Søndergaard J. Drivers for successful long-term lifestyle change, the role of e-health: a qualitative interview study. BMJ Open 2018 Mar 12;8(3):e017466 [FREE Full text] [doi: 10.1136/bmjopen-2017-017466] [Medline: 29530904]
14. Brandt CJ, Søgaard GI, Clemensen J, Sndergaard J, Nielsen JB. General practitioners' perspective on eHealth and lifestyle change: qualitative interview study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e88 [FREE Full text] [doi: 10.2196/mhealth.8988] [Medline: 29666045]
15. Brandt CJ, Søgaard GI, Clemensen J, Søndergaard J, Nielsen JB. Determinants of successful eHealth coaching for consumer lifestyle changes: qualitative interview study among health care professionals. J Med Internet Res 2018 Jul 05;20(7):e237 [FREE Full text] [doi: 10.2196/jmir.9791] [Medline: 29980496]
16. Haste A, Adamson AJ, McColl E, Araujo-Soares V, Bell R. Web-based weight loss intervention for men with type 2 diabetes: pilot randomized controlled trial. JMIR Diabetes 2017 Jul 07;2(2):e14 [FREE Full text] [doi: 10.2196/diabetes.7430] [Medline: 30291100]
17. European Health Examination Survey. National Institute for Health and Welfare. URL: http://www.ehes.info/manuals.htm [accessed 2022-08-21]
18. The Copenhagen City Heart Study Group. The Copenhagen City Heart Study. Osterbroundersøgelsen. A book of tables with data from the first examination (1976-78) and a five year follow-up (1981-83). Scand J Soc Med Suppl 1989;41:1-160. [Medline: 2711133]
19. Understanding A1C Diagnosis. American Diabetes Association. URL: https://www.diabetes.org/a1c/diagnosis [accessed 2022-08-21]
20. 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]
21. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
22. Ryan P. Integrated Theory of Health Behavior Change: background and intervention development. Clin Nurse Spec 2009;23(3):161-70; quiz 171 [FREE Full text] [doi: 10.1097/NUR.0b013e3181a42373] [Medline: 19395894]
23. Magkos F, Fraterrigo G, Yoshino J, Luecking C, Kirbach K, Kelly SC, et al. Effects of moderate and subsequent progressive weight loss on metabolic function and adipose tissue biology in humans with obesity. Cell Metab 2016 Apr 12;23(4):591-601 [FREE Full text] [doi: 10.1016/j.cmet.2016.02.005] [Medline: 26916363]
24. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: A systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
25. Hamaya R, Fukuda H, Takebayashi M, Mori M, Matsushima R, Nakano K, et al. Effects of an mHealth app (Kencom) with integrated functions for healthy lifestyles on physical activity levels and cardiovascular risk biomarkers: observational study of 12,602 users. J Med Internet Res 2021 Apr 26;23(4):e21622 [FREE Full text] [doi: 10.2196/21622] [Medline: 33900203]
26. Cai X, Qiu S, Luo D, Wang L, Lu Y, Li M. Mobile application interventions and weight loss in type 2 diabetes: a meta-analysis. Obesity (Silver Spring) 2020 Mar;28(3):502-509. [doi: 10.1002/oby.22715] [Medline: 31965748]

27. Perri MG, Shankar MN, Daniels MJ, Durning PE, Ross KM, Limacher MC, et al. Effect of telehealth extended care for maintenance of weight loss in rural US communities: a randomized clinical trial. JAMA Netw Open 2020 Jun 01;3(6):e206764 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.6764] [Medline: 32539150]
28. Eberle C, Stichling S. Clinical improvements by telemedicine interventions managing type 1 and type 2 diabetes: systematic meta-review. J Med Internet Res 2021 Feb 19;23(2):e23244 [FREE Full text] [doi: 10.2196/23244] [Medline: 33605889]
29. Silberman JM, Kaur M, Sletteland J, Venkatesan A. Outcomes in a digital weight management intervention with one-on-one health coaching. PLoS One 2020;15(4):e0232221 [FREE Full text] [doi: 10.1371/journal.pone.0232221] [Medline: 32353035]
30. Idris I, Hampton J, Moncrieff F, Whitman M. Effectiveness of a digital lifestyle change program in obese and type 2 diabetes populations: service evaluation of real-world data. JMIR Diabetes 2020 Jan 20;5(1):e15189. [doi: 10.2196/15189] [Medline: 31958064]
31. Lau Y, Chee DGH, Chow XP, Cheng LJ, Wong SN. Personalised eHealth interventions in adults with overweight and obesity: A systematic review and meta-analysis of randomised controlled trials. Prev Med 2020 Mar;132:106001. [doi: 10.1016/j.ypmed.2020.106001] [Medline: 31991155]
32. Islam MM, Poly TN, Walther BA, Jack Li Y. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039 [FREE Full text] [doi: 10.2196/17039] [Medline: 32706724]

## Abbreviations

**BCT:** behavior change technique
**BW:** body weight
**DPP4:** dipeptidyl peptidase-4 inhibitor
**eHealth:** electronic health
**EQ-5D-5L:** European Quality of Life-5 Dimensions
**GP:** general practitioner
**HbA$_{1c}$:** hemoglobin A$_{1c}$
**HCP:** health care professional
**HDL-C:** high-density lipoprotein cholesterol
**LDL-C:** low-density lipoprotein cholesterol
**mHealth:** mobile health
**NCD:** noncommunicable disease
**RCT:** randomized controlled trial
**SMART:** specific, measurable, attainable, relevant, timely
**SWEMWBS:** Short-Warwick-Edinburgh Mental Well-being Scale
**T2D:** type 2 diabetes
**TG:** triglyceride

---

*Edited by G Eysenbach; submitted 23.05.22; peer-reviewed by N Brosnahan; comments to author 14.06.22; revised version received 06.07.22; accepted 31.07.22; published 23.09.22.*

*Please cite as:*
*Hesseldal L, Christensen JR, Olesen TB, Olsen MH, Jakobsen PR, Laursen DH, Lauridsen JT, Nielsen JB, Søndergaard J, Brandt CJ*
*Long-term Weight Loss in a Primary Care–Anchored eHealth Lifestyle Coaching Program: Randomized Controlled Trial*
*J Med Internet Res 2022;24(9):e39741*
*URL: https://www.jmir.org/2022/9/e39741*
*doi:10.2196/39741*
*PMID:36149735*

Original Paper

# A Web-Delivered, Clinician-Led Group Exercise Intervention for Older Adults With Type 2 Diabetes: Single-Arm Pre-Post Intervention

Morwenna Kirwan[1], PhD; Christine L Chiu[1], PhD; Thomas Laing[2], AEP; Noureen Chowdhury[2], MPH; Kylie Gwynne[1], PhD

[1]Faculty of Medicine, Health and Human Sciences, Macquarie University, Macquarie Park, Australia
[2]Diabetes NSW & ACT, Glebe, Australia

**Corresponding Author:**
Morwenna Kirwan, PhD
Faculty of Medicine, Health and Human Sciences
Macquarie University
75 Talavera Road
Macquarie Park
Australia
Phone: 61 0428999072
Email: morwenna.kirwan@mq.edu.au

## Abstract

**Background:** The COVID-19 pandemic created unprecedented shifts in the way health programs and services are delivered. A national lockdown to prevent the spread of COVID-19 in Australia was introduced in March 2020. This lockdown included the closure of exercise clinics, fitness centers, and other community spaces, which, before the pandemic, were used to deliver Beat It. Beat It is an 8-week in-person, community-based, and clinician-led group exercise and education program for adults self-managing diabetes. To continue offering Beat It, it was adapted from an in-person program to a fully web-based supervised group exercise program for adults with type 2 diabetes (T2DM).

**Objective:** This study aims to assess whether the *Beat It Online* program produced comparable health outcomes to the *Beat It* in-person program in terms of improving physical fitness (muscular strength and power, aerobic endurance, balance, and flexibility) and waist circumference in older adults with T2DM.

**Methods:** Australians with T2DM who were aged ≥60 years were included. They were enrolled in *Beat It Online*, a twice-weekly supervised group exercise and education program conducted via videoconference over 8 weeks. Anthropometric measurements and physical fitness parameters were assessed at baseline and completion. The adaptations to *Beat It* are reported using the Model for Adaptation, Design, and Impact, including the type of changes (what, where, when, and for whom), the criteria for making those changes (why and how), and the intended and unintended outcomes. The intended outcomes were comparable functional fitness as well as physical and mental health improvements across demographics and socioeconomic status.

**Results:** A total of 171 adults (mean 71, SD 5.6 years; n=54, 31.6% male) with T2DM were included in the study, with 40.4% (n=69) residing in lower socioeconomic areas. On the completion of the 8-week program, significant improvements in waist circumference, aerobic capacity, muscular strength, flexibility, and balance were observed in both male and female participants (all $P<.001$). The Model for Adaptation, Design, and Impact reports on 9 clinical, practical, and technical aspects of *Beat It* that were adapted for web-based delivery.

**Conclusions:** This study found that *Beat It Online* was just as effective as the in-person program. This adapted program produced comparable health benefits across demographics and socioeconomic status. This study offers important findings for practitioners and policy makers seeking to maintain independence of older people with T2DM, reversing frailty and maximizing functional and physical fitness, while improving overall quality of life. *Beat It Online* offers a flexible and inclusive solution with significant physical and mental health benefits to individuals. Further evaluation of *Beat It* (both in-person and Online) adapted for culturally and linguistically diverse communities will provide greater insights into the efficacy of this promising program.

***(J Med Internet Res 2022;24(9):e39800)*** doi:10.2196/39800

**KEYWORDS**

exercise; fitness; diabetic; physical activity; diabetes; functional fitness; community-based; older adult; videoconference; online; web-based; elder; geriatric; gerontology; balance; movement; internet-based; eHealth; digital health; weight; patient education; translational

## Introduction

According to the International Diabetes Federation, more than half a billion people worldwide have diabetes [1], with type 2 diabetes mellitus (T2DM) accounting for 90% to 95% of cases [1,2]. The upward trajectory of diabetes prevalence has been described as a "pandemic with unprecedented magnitude spiraling out of control" [1].

Physical activity is a cornerstone of T2DM management, along with dietary and pharmacological interventions [3]. Current guidelines recommend that older adults with T2DM engage in at least 150 minutes per week of moderate-intensity aerobic exercise and perform moderate- or high-intensity resistance training on 2 or more days of the week [3]. Currently, two-thirds of older Australians do not meet these guidelines (irrespective of diabetes status) and are deemed insufficiently active to receive physical and mental health benefits [4]. Research has found that older adults who spend less time being physically active have lower overall functional fitness [5]. Functional fitness represents the physical capacity required to carry out typical daily activities independently, without the early onset of fatigue [6,7]. The typical decline in functional fitness with aging is known to lead to frailty, disability, morbidity, and mortality [7].

Physical activity has a substantial role in maintaining functional fitness [8]. Exercise, a structured form of physical activity, has been shown to increase physical fitness—strength, endurance, agility, balance, and flexibility—while also producing other physiological, psychological, and cognitive function benefits for adults with T2DM [3]. Physical fitness provides a protective factor for older adults with T2DM and reduces cardiovascular mortality, all-cause mortality, and the risk of falling [9-11]. Additionally, good physical fitness is known to extend years of active independent living while reducing disability and improving the quality of life [12]. Effective approaches to help older adults with T2DM improve physical fitness, which in turn will maintain functional fitness, are urgently needed.

Small and large-scale community-based supervised group exercise programs for older adults with T2DM have demonstrated effectiveness in improving physical fitness in the short term (immediately after intervention) [13-19], with follow-up studies demonstrating that these improvements can be maintained for up to 1 year [20-22]. Public health recommendations and governmental measures during the ongoing COVID-19 pandemic resulted in various restrictions on daily living including social distancing, isolation, and house confinement. Although these restrictions helped to abate the rate of infection [23], such limitations created considerable disruption to the routine supportive care of those managing T2DM [24]. Many health programs and providers had to pivot to delivering services over the web during this time. A national lockdown to prevent the spread of COVID-19 in Australia was introduced in March 2020. This lockdown included the closure of exercise clinics, fitness centers, and other community spaces, which, before the pandemic, were used to deliver *Beat It*, an 8-week in-person, community-based, and clinician-led group exercise and education program for adults self-managing diabetes. To continue offering this translational program, *Beat It* was adapted from an in-person program to a fully web-based supervised group exercise program for adults with T2DM.

The short- and long-term health and physical fitness outcomes from the traditionally delivered in-person *Beat It* program have been published elsewhere [19,22]. This research found that completing the *Beat It* program significantly improved physical fitness (muscular strength and power, aerobic endurance, balance, and flexibility) and reduced waist circumference for older adults (age ≥60 years), and these benefits were maintained for up to 12 months postprogram [22]. What is unclear is whether the adapted *Beat ItOnline* program provides comparable benefits to participants. Thus, the aim of this study was to assess whether the *Beat It Online* program was effective in improving physical fitness (muscular strength and power, aerobic endurance, balance, and flexibility) and waist circumference during the height of the COVID-19 pandemic.

## Methods

*Beat It Online* is an 8-week program that involves twice-weekly synchronous group exercise and education sessions, delivered via videoconferencing, by accredited exercise physiologists (AEPs).

### Program Design

Programs were delivered by AEPs to participants who resided in New South Wales (NSW) and the Australian Capital Territory (ACT). Exercise sessions include moderate-intensity aerobic, resistance, flexibility, and balance-based exercises, and the education sessions focus on different areas of diabetes self-management. Participants completed a pre-exercise screening, baseline measures, and fitness testing during a web-based 1-on-1 initial consultation with the AEP. This consultation also included motivational interviewing to ascertain participants' goals and their facilitators and barriers to change [25].

The information gathered during the consultation was used to create a customized exercise program with the participant, considering their physical ability, existing fitness levels, and any comorbidities or injuries that could be exacerbated by exercise. Each exercise program comprised home exercise options that participants could replicate with minimal equipment or using items available around the home, including dynamic warm-up and cooldown, and aerobic (eg, walking, stepping, aerobics, and shadow boxing), resistance (eg, body weight, resistance bands, free weights, or household items of equivalent weight—for example, cans of food or bags of rice as hand weights), balance (eg, different standing balance variations), and flexibility exercises (eg, static stretching of major muscle groups). During the web-based group exercise sessions, participants completed their exercise program under the supervision of an AEP, who was able to modify exercises in accordance with the participant's progress.

## Platform and Delivery

Zoom videoconferencing software (version 5.2.1; Zoom Video Communications) was the preferred platform for delivering the group exercise and education sessions. All AEPs purchased a subscription to Zoom as a security precaution to ensure that 2-factor authentication, passcodes, and waiting rooms were available; thus, incidents such as *Zoom bombing* and other security concerns were prevented.

## Facilitator Training

All AEPs completed a specialized facilitator training program called *Beat It Trainer* to ensure the consistent and effective delivery of the program. This accredited continuing professional development training course, certified by *Diabetes Qualified*, a subsidiary of Diabetes NSW & ACT, consisted of 12 hours of asynchronous web-based modules and assessments, followed by a 1-day synchronous session focused on the clinical, practical, and technical requirements of delivering the *Beat It* program. Additionally, AEPs also completed a supplemental 2-hour synchronous training session focused on considerations for web-based delivery. To maintain this certification, all AEPs are required to complete a refresher course within 2 years of completing the initial training.

## Recruitment

From May 2020 to October 2021, participants were recruited through email using the National Diabetes Services Scheme database or advertisements on the Diabetes NSW & ACT website. Prior to commencing, participants were required to provide evidence of medical clearance to exercise from their general practitioner and were then eligible to attend their web-based initial health and fitness assessment with their AEP. Standard exclusion criteria for the *Beat It Program* (in-person and Online) are outlined in Multimedia Appendix 1. Health and fitness assessments were performed at baseline and at the completion of the program. For inclusion into the current study, participants had to be aged ≥60 years, have a clinical diagnosis of T2DM, and have completed the web-based health and fitness assessment at baseline and at 8 weeks at the completion of the program.

## Ethics Approval

This study was approved by the Macquarie University Human Ethics Committee (5201950887424).

## Study Design and Measures

This study used a pre-post evaluation design where participants completed web-based physical assessment sessions at baseline and at 8 weeks of *Beat It Online*, as well as an evaluation questionnaire prior to and at the completion of the program. Sociodemographic information, including gender, date of birth, and residential postcode, was collected. Postcodes were used as a measure of socioeconomic status as previously described [19]. Height and weight were used to calculate BMI, and BMI was then categorized as healthy, overweight, and obese, according to the World Health Organization [26]. An assessment of upper and lower body muscular strength, aerobic capacity, flexibility, and balance were performed using the 30-second sit-to-stand test; 30-second seated arm curl test; 2-minute step test; seated sit-and-reach test; and 1-legged stand test, respectively [7]. The average of left and right sit-and-reach test and 1-legged stand test was used. Participants with missing data for gender, age, and waist circumference measures were excluded from analysis.

The Diabetes Empowerment Scale (DES) is a measure of psychosocial self-efficacy that was developed specifically for a diabetes population. The DES short form contains 8 items that can be used to provide a brief general assessment of diabetes psychosocial self-efficacy [27]. Participants are asked to indicate their agreement to 8 individual statements related to their ability to manage diabetes. Each item is rated on a 5-point scale: 5=strongly agree, 4=somewhat agree, 3=neutral, 2=somewhat disagree, and 1=strongly disagree. The overall score is the sum of the scores for each individual item, divided by the number of items (8 items).

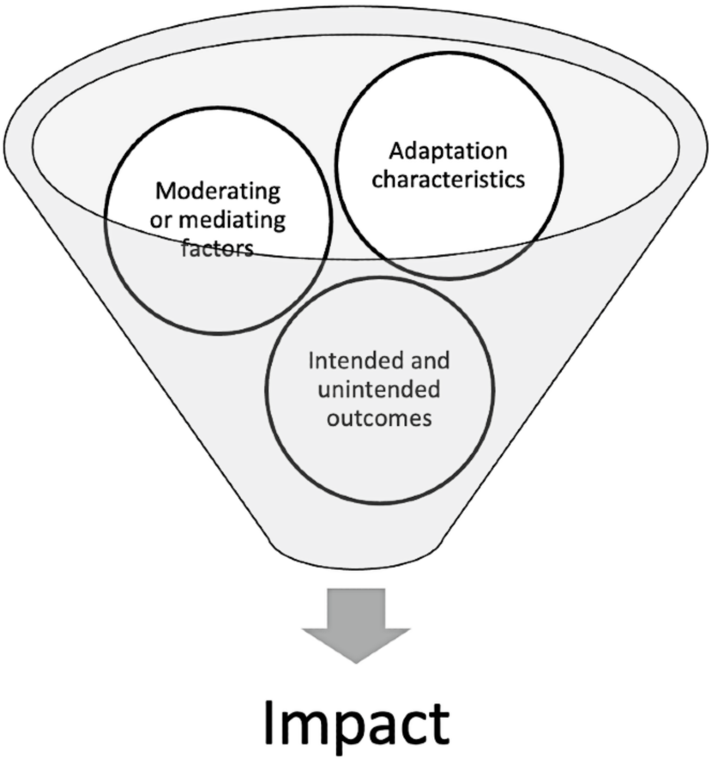
The Patient Activation Measure (PAM) assesses a person's self-reported knowledge, skill, and confidence for the self-management of their health or chronic condition [28]. Each item is rated on 4-point scale: from 1=strongly disagree to 4=strongly agree, with an additional "not applicable" option. The total PAM score is the sum of raw scores for each item, divided by the number of items excepting the nonapplicable items and then multiplied by 13. Higher scores are associated with the person taking action.

## Analysis

Data analysis was performed using SPSS statistical software (version 27; IBM Corp). Mean and SD were calculated for continuous variables. Frequencies and percentages were calculated for categorical variables, excluding participants with missing data for that variable. Paired 2-tailed $t$ tests were performed to identify differences in fitness measures at baseline and program completion.

The adaptation of *Beat It* has been represented using the Model for Adaptation, Design, and Impact (MADI) [29]. This model has 3 domains: adaption characteristics (what was modified and how, for whom and when, and who was involved in decision making); moderating or mediating factors (adaptation aligned with the core elements of the intervention or implementation strategy, clear goals for adaptation, and adaptations implemented consistently); and intended and unintended outcomes. The analysis of the 3 domains together helps explain impact (Figure 1).

**Figure 1.** Model for Adaptation, Design, and Impact.



## Results

### Beat It MADI Results

The *Beat It Online* adaptations using the MADI model are shown in Table 1. There were 9 clinical, practical, and technical aspects of *Beat It* that were adapted for web-based delivery. Each of these aspects had moderating and mediating factors to support program fidelity while operationalizing *Beat It* in a web-based context. The intended outcomes were comparable functional fitness and physical and mental health improvements across demographics and socioeconomic status. These outcomes are reported in Table 2. At this stage, unintended outcomes are unknown and will be explored in a follow-up study. The impact was that *Beat It Online* was as effective as the in-person program.

**Table 1.** Model for Adaptation, Design, and Impact report of adaptations and mediating and moderating factors.

| Adaptation areas | Mediating and moderating factors, program stage | |
|---|---|---|
| | Beat It (in-person) | Beat It Online |
| Facilitator training | 12 hours web-based learning and 1-day in-person practical training | Additional 2 hours covering key considerations for web-based delivery |
| Marketing | Direct mail (via post), email, and website | Digital only—email and website |
| Participant resources | *Beat It* participant handbook, home exercise resource, and Theraband | Additional resources guiding participants on technical requirements and equipment required to access web-based sessions:<br>• internet connectivity and speed tests<br>• how to set-up or log in to videoconferencing platform<br>• participant pre–web-based session checklist outlining key safety considerations to be followed every session |
| Trainer resources | *Beat It* in-person delivery manual and *Beat It* facilitator manual (education sessions) | Additional content added to the *Beat It Online* delivery manual:<br>• assessing participant suitability<br>• additional Beat It Trainer requirements<br>• required equipment and technical capabilities, safety requirements, legal and professional practice requirements, and privacy and record keeping considerations<br>• web-based delivery considerations specific to exercise and education session delivery<br>• assisting Beat It trainers in selecting appropriate videoconferencing platforms<br>• ways to maintain communication and establish visual cues (eg, Zoom gallery view) to ensure participants are performing exercises confidently and safely<br>• trainer checklists and session guides to follow prior, during, and following each session |
| Medical clearance | Standardized medical clearance form, including recommended program inclusion or exclusion criteria, medical history, medications, and latest Hemoglobin $A_{1c}$ and lipid test results; participants typically bring a physical copy of the medical clearance form to initial consultation with a *Beat It* trainer | Additional considerations and exclusion criteria for determining suitability to join the web-based program, including:<br>• client digital literacy<br>• client needs and goals<br>• risks including precautions and contraindications<br>• physical capacity of client to undertake session<br>• client ability to provide consent<br>• capacity to access technology<br>• need and availability for a client support person (eg, family, carer, and allied health assistant) to assist in consult and sessions<br>• further information relating to hypoglycemia frequency and falls risk requested from referring medical practitioner<br>• additional instructions relating to sending medical clearance information safely and securely via appropriate web-based methods (eg, encrypted email and fax) to the participant's Beat It trainer |
| Preprogram | Preprogram resources sent including welcome letter confirming program registration, medical clearance, and initial consultation process and the *Beat It* trainer books initial assessment appointment | Additional resources guiding participants on what equipment is required to access *Beat It Online*, including:<br>• technical requirements (eg, internet connection, appropriate device for video calls, and active email address)<br>• initial assessment equipment (eg, measuring tape, weight scales, suitable chair, and hand weight or substitute to perform exercise tests from home)<br>• safety considerations (eg, appropriate exercise space, clothing, and access to blood glucose monitor and hypoglycemia treatment)<br>• provision of step-by-step guide of how to access the Beat It trainer's selected videoconferencing platform<br>• at the time of assessment booking, the Beat It trainer also assessed participant's technical proficiency with using the videoconferencing tool |

| Adaptation areas | Mediating and moderating factors, program stage | |
|---|---|---|
| | Beat It (in-person) | Beat It Online |
| Initial and final assessment | • Conducted in person<br>• Obtain medical clearance, participant informed consent, and emergency contact information and complete prescreening questionnaire<br>• Complete baseline measurements, including height, weight, waist circumference, blood pressure, and heart rate<br>• Complete exercise tests including the 6-minute walk test, 30-second sit-to-stand test, 30-second seated arm curl test, seated sit-and-reach test, and 1-legged stand test<br>• Goal setting | Differences:<br>• both assessments conducted over the web<br>• informed consent form sent digitally<br>• further participant emergency contact information was collected, including physical address participant would be completing web-based exercise sessions from, presence of friend or family at this address during sessions, and education related to having access to blood glucose monitor and hypoglycemia treatment available during sessions<br>• baseline measurement protocols were adapted to support participants complete these themselves with guidance from a Beat It trainer via video (eg, using string or sewers tape to measure waist circumference) and requested from the referring general practitioner as part of medical clearance process (eg, resting blood pressure if participant does not have access to a blood pressure monitor)<br>• exercise testing protocols were modified to allow participant completion from home<br>• 6-minute walk test was replaced with a 2-minute step-in-place test to account for space constraints and allow participants to be monitored on camera<br>• household items were used to facilitate other tests (eg, 2-3 kg bag of rice for seated arm curl test and ruler used for seated sit-and-reach test) |
| Exercise sessions | Capped at 12 participants per session; in-person exercise sessions consist of a warm-up, followed by a combination of aerobic, resistance, balance, and flexibility exercises tailored to participants abilities, followed by a cooldown period | Differences:<br>• capped at 6 participants per session to enable adequate supervision in a web-based setting<br>• corresponding pre-exercise checklists for the Beat It trainer and participants detailing important steps the participants must take leading into each exercise session including ensuring technology is setup correctly allowing the sessions to be viewed clearly, confirming participants have hypoglycemia treatment available, ensuring their exercise area is free from obstructions, and asking participants to take pre- or postexercise blood glucose measurements<br>• guidelines on camera and microphone settings to ensure Beat It trainers can be seen and heard, and participants can be monitored effectively and the use of visual cues and telehealth functions to provide feedback and breaking up exercises to check-in with participants<br>• Beat It trainers to structure sessions by providing 3-4 different options for each exercise delivered to the group; these options include regressions and progressions for each exercise and ensure participants can complete a similar exercise at the same time, dependent on their ability<br>• Beat It trainers encouraged to keep sessions as creative, fun, and engaging as possible with ideas such as dress-up themes, activity-based challenges, and games to encourage social interaction among the group |
| Education sessions | 6 x 30 min person-centered education sessions on various lifestyle and diabetes management topics delivered in person | Differences:<br>• delivered over the web<br>• used screen-sharing functions and web-based whiteboards to collate participant responses |

**Table 2.** Summary of assessment data.

| Assessment | Male | | | | Female | | | |
|---|---|---|---|---|---|---|---|---|
| | n | Baseline, mean (SD) | Postprogram, mean (SD) | *P* value | n | Baseline, mean (SD) | Postprogram, mean (SD) | *P* value |
| Waist circumference (cm) | 54 | 114.3 (12.5) | 110.6 (12.5) | <.001 | 117 | 102.7 (14.3) | 100 (13.3) | <.001 |
| Seated sit-and-reach (cm) | 46 | –7.8 (11.1) | –4.3 (10.2) | <.001 | 106 | –2.5 (11.9) | 0.2 (11.8) | <.001 |
| 30-second sit-to-stand (reps) | 51 | 13.2 (3.9) | 16.2 (5.2) | <.001 | 116 | 12.6 (4.6) | 15.7 (5.3) | <.001 |
| 1-legged stand test (s) | 52 | 27.1 (20.3) | 34.5 (18.5) | <.001 | 115 | 25.4 (19.9) | 33.2 (20.5) | <.001 |
| 2-minute step test (reps) | 45 | 68.5 (22.5) | 87.9 (26.8) | <.001 | 105 | 69.2 (26.2) | 86.8 (30.3) | <.001 |
| Arm curl (reps) | 50 | 20.1 (9) | 26.0 (8.8) | <.001 | 114 | 18.9 (8.6) | 23.5 (8.5) | <.001 |

### Beat It Online Participant Results

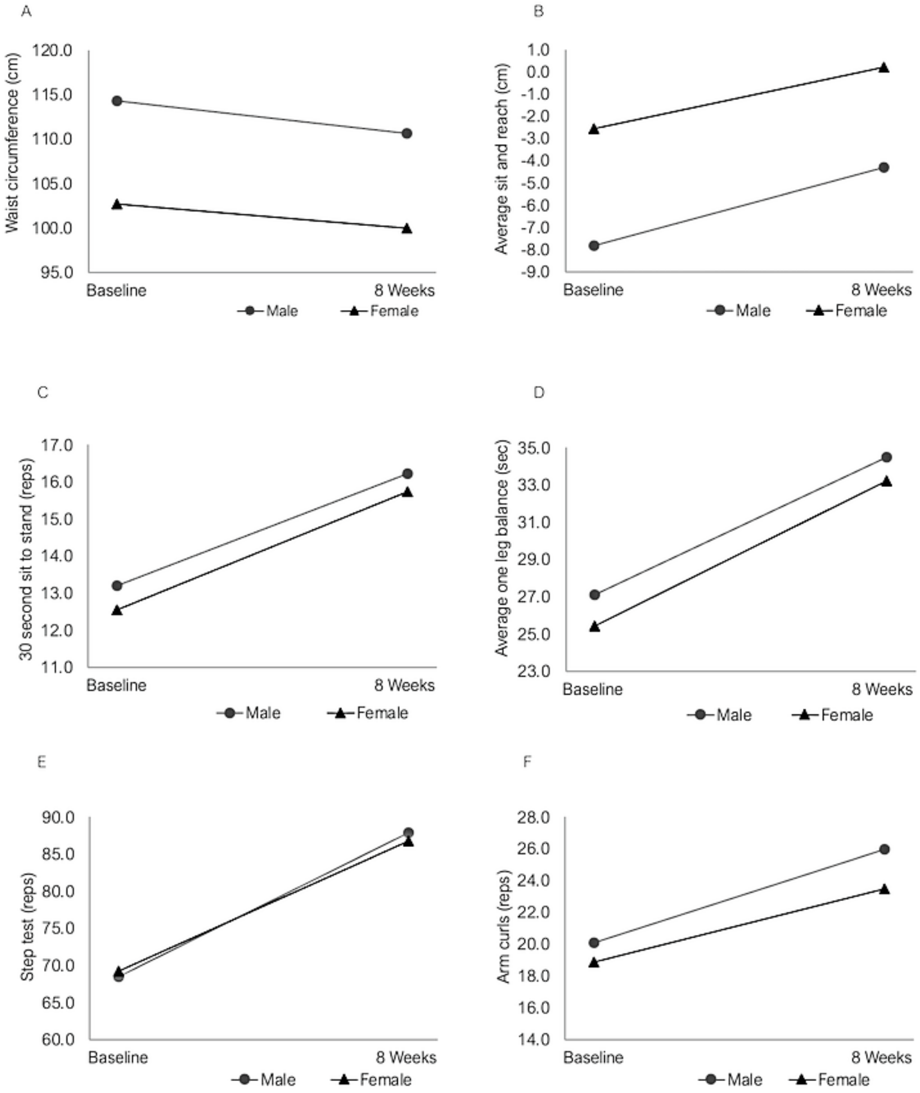
A total of 171 individuals were included in the study. These individuals were aged ≥60 years, had reported a diagnosis of T2DM, and had participated in *Beat It Online*. Of the 171 individuals, 54 (32%) were male, age ranged from 60 to 89 years with a mean age of 71 (SD 5.6) years, and 69 (40%) resided in lower socioeconomic areas. Participants attended between 5 and 16 exercise sessions, with 70% (n=119) of the cohort attending at least 14 out of the 16 sessions. For the education sessions, participants attended between 0 and 6 sessions, with 80% (n=137) of the cohort attending at least 5 out of the 6 education sessions.

Improvements in waist circumference, aerobic capacity, muscular strength, flexibility, and balance were observed postprogram in both male and female participants (Table 2 and Figure 2).

Survey evaluation data was received from 49 (29%) of the 171 participants. Most participants rated their health as being good to excellent both at baseline and postprogram (n=38, 78% vs n=36, 74%), whereas improvements in their general quality of life rated as being good to excellent were reported postprogram (n=38, 78% vs n=47, 96%).

Significant improvements in DES scores were reported postprogram (3.9 vs 4.3; *P*<.001), whereas nonsignificant improvements in PAM scores were observed (44.3 vs 45.4; *P*=.07).

**Figure 2.** Mean fitness and health measures at baseline and at program completion, stratified by gender.



## Discussion

### Principal Findings

This study found that *Beat It Online* was just as effective as the in-person program [19,22]. This adapted program produced comparable health benefits, across demographics and socioeconomic status. The COVID-19 pandemic has necessitated limitations on physical interaction, requiring the rapid web-based adaptation of key health programs and services. This study found that the *Beat It Online* program, an 8-week clinician-led supervised group exercise program, significantly improved physical fitness (aerobic capacity, strength, balance, and flexibility), reduced waist circumference, and improved self-efficacy and quality of life in older adults with T2DM. This finding is important because globally, it has been reported that during the home confinement of lockdowns and necessary isolation, most individuals decreased their physical activity levels and increased sedentary behavior [30], and those with T2DM had increased body weight and worsened their glycemic control [31].

In this study, we used MADI, an implementation science framework, to transparently evaluate and report on the adaptations to the *Beat It* program [29]. We identified important adaptation characteristics, their intended and unintended impacts, and potential mediators and moderators of adaptations' impact on outcomes. The adjustments made to the processes and protocols for *Beat It* were similar to that outlined by Schwartz and colleagues [32] in their pivot to delivering web-based, synchronous group exercise programs to older adults. It is important to not assume that an exercise specialist (or any health practitioner) who primarily delivers services in-person will be capable of delivering a comparable service over the web. Research has shown that the most prevalent barrier for the organizational adoption of telehealth, by a wide margin, is technically challenged health practitioners [33]. A comprehensive capability framework, such as the one developed by Davies and colleagues [34], provides important guidance on the core capabilities, curricula, and professional development needed in this space.

The focus in diabetes care has traditionally been around the optimization of glycemic control and deterrence of complications [35]. However, the prevention of frailty and improvement in physical function have now emerged as new targets of diabetes management [35]. This focus is mainly driven by the significant adverse impact that early onset frailty and declines in functional fitness have on health outcomes, including glycemic control, independence, and quality of life in people with T2DM [35]. The *Beat It Online* program produced significant improvements in all physical fitness measures, as well as self-efficacy and self-reported quality of life, in a cohort that was aged 71 (SD 5.6; range 60-89) years. At a time when older adults across the world were told to isolate at home, *Beat It Online* provided a socially supportive setting for vulnerable individuals with a chronic condition and comorbidities to exercise safely, boost their mood, and connect with their peers while following health directives related to COVID-19. This program is particularly important given that research has shown that stay-at-home orders have resulted in the loss of formal and incidental social connections, putting older adults at acute risk for loneliness [36]. Furthermore, participation in moderate to high volumes of physical activity during and following periods of COVID-19 containment has been associated with better mental health and well-being compared to inactive adults [37].

The *Beat It Online* program has debunked the pervasive stereotype that older adults are resistant and incapable of engaging with new technologies [38]. The participants in this study demonstrated that they were skilled at using videoconferencing software in a way that helped them improve their physical health and mental well-being. This openness to using technology for a clear purpose is well supported in the literature [39] and aligns directly with research showing that most older adults have a strong motivation to learn new skills and continue living fully through learning [40]. This finding has implications for practitioners and policy makers who may assume that this demographic only prefers in-person programs and services.

A limitation of this study is that it used a pre-post evaluation with no comparison group, which is a common design for translational community-based programs [41]. Another limitation is that we only evaluated the short-term effectiveness of the *Beat It Online* program. A longer follow-up is needed to ascertain if participants maintained the benefits gained from this program, any unintended outcomes, and whether *Beat It Online* is pragmatic outside of a global pandemic. We conducted a 1-year follow-up of the in-person *Beat It* program and concluded that participants maintained improvements in their health 12 months after completing the 8-week program [22].

## Conclusions

This study revealed that a fully web-delivered, clinician-led, and supervised group exercise program provided important health benefits to older adults with T2DM. This study offers important findings for practitioners and policy makers seeking to maintain independence of older persons with T2DM, reversing frailty and maximizing functional and physical fitness while improving overall quality of life. The COVID-19 pandemic has created unprecedented shifts in the way key health programs and services are delivered. *Beat It Online* offers a flexible and inclusive solution with significant physical and mental health benefits to individuals. Further evaluation of *Beat It* (both in-person and Online) adapted for culturally and linguistically diverse communities will provide greater insights into the efficacy of this promising program.

## Acknowledgments

The *Beat It* program delivery costs were funded by the National Diabetes Services Scheme (NDSS). The NDSS is an Australian Government Initiative administered by Diabetes Australia. The authors received no funding to conduct this study. The NDSS had no role in the outcomes reported in this study. Diabetes NSW & ACT cofunded the publication costs for the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Standard exclusion criteria for Beat It (in-person) and Beat It Online.

[DOCX File , 16 KB - jmir_v24i9e39800_app1.docx ]

## References

1. International Diabetes Federation. IDF Diabetes Atlas Tenth Edition. IDF Diabetes Atlas. 2021. URL: https://diabetesatlas.org/ [accessed 2022-09-08]
2. American Diabetes Association Professional Practice Committee. Classification and diagnosis of diabetes: standards of medical care in diabetes-2022. Diabetes Care 2022 Jan 01;45(Suppl 1):S17-S38 [FREE Full text] [doi: 10.2337/dc22-S002] [Medline: 34964875]
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 Feb 01;54(2):353-368. [doi: 10.1249/MSS.0000000000002800] [Medline: 35029593]
4. Insufficient physical activity. Australian Institute of Health and Welfare. 2020 Oct 20. URL: https://www.aihw.gov.au/reports/risk-factors/insufficient-physical-activity/contents/insufficient-physical-activity [accessed 2022-04-01]
5. Santos DA, Silva AM, Baptista F, Santos R, Vale S, Mota J, et al. Sedentary behavior and physical activity are independently related to functional fitness in older adults. Exp Gerontol 2012 Dec;47(12):908-912 [FREE Full text] [doi: 10.1016/j.exger.2012.07.011] [Medline: 22884978]
6. Kostić R, Uzunović S, Pantelić S, Đuraškovič R. A comparative analysis of the indicators of the functional fitness of the elderly. Facta universitatis-series: physical education and sport 2011 Aug 4;9(2):161-171 [FREE Full text]
7. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist 2013 Apr;53(2):255-267. [doi: 10.1093/geront/gns071] [Medline: 22613940]
8. Milanović Z, Pantelić S, Trajković N, Sporiš G, Kostić R, James N. Age-related decrease in physical activity and functional fitness among elderly men and women. Clin Interv Aging 2013 May 21;8:549-556 [FREE Full text] [doi: 10.2147/CIA.S44112] [Medline: 23723694]
9. Bennett WL, Ouyang P, Wu AW, Barone BB, Stewart KJ. Fatness and fitness: how do they influence health-related quality of life in type 2 diabetes mellitus? Health Qual Life Outcomes 2008 Dec 04;6:100 [FREE Full text] [doi: 10.1186/1477-7525-6-110] [Medline: 19055828]
10. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. JAMA 2009 May 20;301(19):2024-2035. [doi: 10.1001/jama.2009.681] [Medline: 19454641]
11. Crews RT, Yalla SV, Fleischer AE, Wu SC. A growing troubling triad: diabetes, aging, and falls. J Aging Res 2013 Feb 16;2013:342650 [FREE Full text] [doi: 10.1155/2013/342650] [Medline: 23476773]
12. Sun F, Norman IJ, While AE. Physical activity in older people: a systematic review. BMC Public Health 2013 May 06;13:449 [FREE Full text] [doi: 10.1186/1471-2458-13-449] [Medline: 23648225]
13. Mendes R, Sousa N, Reis VM, Themudo-Barata JL. Implementing low-cost, community-based exercise programs for middle-aged and older patients with type 2 diabetes: what are the benefits for glycemic control and cardiovascular risk? Int J Environ Res Public Health 2017 Sep 13;14(9):1057 [FREE Full text] [doi: 10.3390/ijerph14091057] [Medline: 28902144]
14. Mendes R, Sousa N, Themudo-Barata J, Reis V. Impact of a community-based exercise programme on physical fitness in middle-aged and older patients with type 2 diabetes. Gac Sanit 2016 May;30(3):215-220 [FREE Full text] [doi: 10.1016/j.gaceta.2016.01.007] [Medline: 26900099]
15. Gallé F, Di Onofrio V, Miele A, Belfiore P, Liguori G. Effects of a community-based exercise and motivational intervention on physical fitness of subjects with type 2 diabetes. Eur J Public Health 2019 Apr 01;29(2):281-286. [doi: 10.1093/eurpub/cky140] [Medline: 30052924]
16. Grelier S, Serresse O, Boudreau-Lariviere C, Zory R. Effects of a three-month combined training program on the cardiopulmonary and muscle strength capacities of type 2 diabetic subjects. J Sports Med Phys Fitness 2013 Feb;53(1):56-64. [Medline: 23470912]
17. Castaneda C, Layne JE, Munoz-Orians L, Gordon PL, Walsmith J, Foldvari M, et al. A randomized controlled trial of resistance exercise training to improve glycemic control in older adults with type 2 diabetes. Diabetes Care 2002 Dec;25(12):2335-2341. [doi: 10.2337/diacare.25.12.2335] [Medline: 12453982]
18. Dunstan DW, Vulikh E, Owen N, Jolley D, Shaw J, Zimmet P. Community center-based resistance training for the maintenance of glycemic control in adults with type 2 diabetes. Diabetes Care 2006 Dec;29(12):2586-2591. [doi: 10.2337/dc06-1310] [Medline: 17130189]
19. Kirwan M, Chiu CL, Hay M, Laing T. Community-based exercise and lifestyle program improves health outcomes in older adults with type 2 diabetes. Int J Environ Res Public Health 2021 Jun 07;18(11):6147 [FREE Full text] [doi: 10.3390/ijerph18116147] [Medline: 34200198]
20. Gallè F, Krakauer JC, Krakauer NY, Valerio G, Liguori G. Can an exercise-based educational and motivational intervention be durably effective in changing compliance to physical activity and anthropometric risk in people with type 2 diabetes? a

follow-up study. Int J Environ Res Public Health 2019 Feb 27;16(5):701 [FREE Full text] [doi: 10.3390/ijerph16050701] [Medline: 30818773]
21. Mathieu M, Brochu M, Béliveau L. DiabetAction: changes in physical activity practice, fitness, and metabolic syndrome in type 2 diabetic and at-risk individuals. Clin J Sport Med 2008 Jan;18(1):70-75. [doi: 10.1097/JSM.0b013e31815c1d68] [Medline: 18185042]
22. Kirwan M, Gwynne K, Laing T, Hay M, Chowdhury N, Chiu CL. Can health improvements from a community-based exercise and lifestyle program for older adults with type 2 diabetes be maintained? a follow up study. Diabetology 2022 May 19;3(2):348-354. [doi: 10.3390/diabetology3020025]
23. Hossain MM, Sultana A, Purohit N. Mental health outcomes of quarantine and isolation for infection prevention: a systematic umbrella review of the global evidence. Epidemiol Health 2020 Jun 02;42:e2020038 [FREE Full text] [doi: 10.4178/epih.e2020038] [Medline: 32512661]
24. Carr MJ, Wright AK, Leelarathna L, Thabit H, Milne N, Kanumilli N, et al. Impact of COVID-19 restrictions on diabetes health checks and prescribing for people with type 2 diabetes: a UK-wide cohort study involving 618 161 people in primary care. BMJ Qual Saf 2022 Jul;31(7):503-514 [FREE Full text] [doi: 10.1136/bmjqs-2021-013613] [Medline: 34642228]
25. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change. 3rd ed. New York, NY: Guilford Press; 2012.
26. A healthy lifestyle - WHO recommendations. World Health Organization. 2010 May 06. URL: http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi [accessed 2022-10-14]
27. Anderson RM, Fitzgerald JT, Gruppen LD, Funnell MM, Oh MS. The Diabetes Empowerment Scale-Short Form (DES-SF). Diabetes Care 2003 May;26(5):1641-1642. [doi: 10.2337/diacare.26.5.1641-a] [Medline: 12716841]
28. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
29. Kirk MA, Moore JE, Wiltsey Stirman S, Birken SA. Towards a comprehensive model for understanding adaptations' impact: the model for adaptation design and impact (MADI). Implement Sci 2020 Jul 20;15(1):56 [FREE Full text] [doi: 10.1186/s13012-020-01021-y] [Medline: 32690104]
30. Ammar A, Brach M, Trabelsi K, Chtourou H, Boukhris O, Masmoudi L, et al. Effects of COVID-19 home confinement on eating behaviour and physical activity: results of the ECLB-COVID19 international online survey. Nutrients 2020 May 28;12(6):1583 [FREE Full text] [doi: 10.3390/nu12061583] [Medline: 32481594]
31. Biamonte E, Pegoraro F, Carrone F, Facchi I, Favacchio G, Lania AG, et al. Weight change and glycemic control in type 2 diabetes patients during COVID-19 pandemic: the lockdown effect. Endocrine 2021 Jun;72(3):604-610 [FREE Full text] [doi: 10.1007/s12020-021-02739-5] [Medline: 33950348]
32. Schwartz H, Har-Nir I, Wenhoda T, Halperin I. Staying physically active during the COVID-19 quarantine: exploring the feasibility of live, online, group training sessions among older adults. Transl Behav Med 2021 Mar 16;11(2):314-322 [FREE Full text] [doi: 10.1093/tbm/ibaa141] [Medline: 33447852]
33. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
34. Davies L, Hinman RS, Russell T, Lawford B, Bennell K, International Videoconferencing Steering Group. An international core capability framework for physiotherapists to deliver quality care via videoconferencing: a Delphi study. J Physiother 2021 Oct;67(4):291-297 [FREE Full text] [doi: 10.1016/j.jphys.2021.09.001] [Medline: 34521586]
35. Ahmad E, Sargeant JA, Yates T, Webb DR, Davies MJ. Type 2 diabetes and impaired physical function: a growing problem. Diabetology 2022 Jan 11;3(1):30-45. [doi: 10.3390/diabetology3010003]
36. Lopez KJ, Tong C, Whate A, Boger J. “It’s a whole new way of doing things”: the digital divide and leisure as resistance in a time of physical distance. World Leis J 2021 Sep 15;63(3):281-300. [doi: 10.1080/16078055.2021.1973553]
37. Faulkner J, O'Brien WJ, Stuart B, Stoner L, Batten J, Wadsworth D, et al. Physical activity, mental health and wellbeing of adults within and during the easing of COVID-19 restrictions, in the United Kingdom and New Zealand. Int J Environ Res Public Health 2022 Feb 04;19(3):1792 [FREE Full text] [doi: 10.3390/ijerph19031792] [Medline: 35162815]
38. Mariano J, Marques S, Ramos MR, Gerardo F, Cunha C, Girenko A, et al. Too old for technology? stereotype threat and technology use by older adults. Behaviour & Information Technology 2021 Feb 07;41(7):1503-1514 [FREE Full text] [doi: 10.1080/0144929X.2021.1882577]
39. Vroman K, Arthanat S, Lysack C. “Who over 65 is online?” older adults’ dispositions toward information communication technology. Comput Human Behav 2015 Feb;43:156-166 [FREE Full text] [doi: 10.1016/j.chb.2014.10.018]
40. Boulton-Lewis GM, Buys L, Lovie-Kitchin J. Learning and active aging. Educ Gerontol 2006 Apr 1;32(4):271-282. [doi: 10.1080/03601270500494030]
41. Titler M. Translation Research in Practice: An Introduction. Online J Issues Nurs 2018 May 31;23(2) [FREE Full text] [doi: 10.3912/OJIN.Vol23No02Man01]

## Abbreviations

**ACT:** Australian Capital Territory
**AEP:** accredited exercise physiologist
**DES:** Diabetes Empowerment Scale
**MADI:** Model for Adaptation, Design, and Impact
**NSW:** New South Wales
**PAM:** Patient Activation Measure
**T2DM:** type 2 diabetes mellitus

---

*Edited by G Eysenbach; submitted 24.05.22; peer-reviewed by L Davies, D Goodrich; comments to author 01.07.22; revised version received 14.07.22; accepted 31.08.22; published 23.09.22.*

*Please cite as:*
*Kirwan M, Chiu CL, Laing T, Chowdhury N, Gwynne K*
*A Web-Delivered, Clinician-Led Group Exercise Intervention for Older Adults With Type 2 Diabetes: Single-Arm Pre-Post Intervention*
*J Med Internet Res 2022;24(9):e39800*
*URL: https://www.jmir.org/2022/9/e39800*
*doi:10.2196/39800*
*PMID:36149745*

Original Paper

# An Internet-Based Education Program for Human Papillomavirus Vaccination Among Female College Students in Mainland China: Application of the Information-Motivation-Behavioral Skills Model in a Cluster Randomized Trial

Mingyu Si[1], MD; Xiaoyou Su[1*], PhD; Yu Jiang[1*], PhD; Wenjun Wang[2], PhD; Xi Zhang[3], PhD; Xiaofen Gu[4], PhD; Li Ma[5], PhD; Jing Li[6], PhD; Shaokai Zhang[7], PhD; Zefang Ren[8], PhD; Yuanli Liu[9], PhD; Youlin Qiao[1], PhD

[1]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[2]School of Nursing, Jining Medical University, Jining, China
[3]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Beijing Office for Cancer Prevention and Control, Peking University Cancer Hospital & Institute, Beijing, China
[4]Affiliated Tumor Hospital, Xinjiang Medical University, Urumqi, China
[5]School of Public Health, Dalian Medical University, Dalian, China
[6]West China School of Public Health, West China Forth Hospital, Sichuan University, Chengdu, China
[7]Henan Cancer Hospital, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China
[8]School of Public Health, Sun Yat-sen University, Guangzhou, China
[9]School of Health Policy and Management, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[*]these authors contributed equally

**Corresponding Author:**
Xiaoyou Su, PhD
School of Population Medicine and Public Health
Chinese Academy of Medical Sciences and Peking Union Medical College
5 DongDanSanTiao, Dongcheng District
Beijing, 100730
China
Phone: 86 10 65256070
Email: suxiaoyou@hotmail.com

## Abstract

**Background:** Patients diagnosed with cervical cancer in the last 2 decades were mainly young females. Human papillomavirus (HPV) vaccination is the most radical way to prevent HPV infection and cervical cancer. However, most female college students in mainland China have not yet been vaccinated, and their relevant knowledge is limited. Theory-based education delivered via the internet is a potentially accessible and useful way to promote HPV vaccination among this population.

**Objective:** This 3-month follow-up study intended to identify the feasibility and efficacy of an information-motivation-behavioral skills (IMB) model–based online intervention for promoting awareness and willingness regarding HPV vaccination among female college students.

**Methods:** A 7-day online HPV education program for female college students in mainland China was developed using a cluster randomized trial design. Recruitment and questionnaire surveys were performed online without face-to-face contact. SPSS 23.0 was used for statistical analysis. The chi-square test and *t* test were used to compare differences in qualitative and continuous variables between intervention and control groups. The generalized estimating equation was used to test the effectiveness of the intervention with a consideration of the time factor.

**Results:** Among 3867 participants, 102 had been vaccinated against HPV before the study (vaccination rate of 2.6%). A total of 3484 participants were followed up after the baseline survey, with no statistical difference in the loss rate between the intervention and control groups during the intervention and follow-up periods. At different follow-up time points, HPV-related knowledge, and the motivation, behavioral skills, and willingness regarding HPV vaccination were higher in the intervention group than in the control group. HPV-related knowledge was statistically different between the 2 groups, while the motivation, behavioral skills, and willingness regarding HPV vaccination only showed statistical differences right after the intervention, reaching a peak right after the intervention and then gradually reducing over time. Furthermore, there was no statistical difference in the HPV vaccination rate between the 2 groups.

**Conclusions:** IMB model–based online education could be a promising way to increase the HPV vaccination rate and reduce the burden of HPV infection and cervical cancer among high-risk female college students in China.

**Trial Registration:** Chinese Clinical Trial Registry ChiCTR1900025476; http://www.chictr.org.cn/showprojen.aspx? proj=42672

**International Registered Report Identifier (IRRID):** RR2-DOI:10.1186/s12889-019-7903-x

***(J Med Internet Res 2022;24(9):e37848)*** doi:10.2196/37848

**KEYWORDS**

human papillomavirus vaccination; internet-based education; information-motivation-behavioral skills model; female college students; China

## *Introduction*

### Background

Cervical cancer, a serious infectious disease mostly caused by high-risk human papillomavirus (HPV) types, is the fourth most common reason worldwide for tumor-related hospital admission in women [1,2]. In 2018, in China, cervical cancer resulted in 48,000 deaths and 106,000 cases [1]. Previous studies found that young women were at higher risk of HPV infection, and showed elevated HPV infection problems owing to their age and the sexually transmitted nature of the virus [3-5]. Female college students, most of whom were ready for sexual debut and had relatively low awareness of disease prevention, similarly demonstrated HPV infection problems right after starting college [4,6].

At the end of 2020, China initiated a program for cancer prevention jointly issued by 10 ministries, including the National Health Commission, and joined in the action of the "global strategy to accelerate the elimination of cervical cancer" advocated by the World Health Organization [7-9]. Since the approval of preventive HPV vaccines in mainland China in 2016, the HPV vaccination rate has remained poor, highlighting the importance of understanding the reasons for vaccination refusal and the ways to promote HPV vaccination among young populations [6,10].

Improving the HPV vaccination rate and vaccination willingness is the most salient way to eliminate cervical cancer [11,12]. Most published studies on the willingness to undergo HPV vaccination among Chinese college students were conducted before the availability of HPV vaccines in China [13-15]. Information about the actual rate of HPV vaccination and willingness to undergo HPV vaccination after the approval of HPV vaccines is scarce. Interventions guided by health behavioral theories were widely applied in previous studies [16-19]. A review conducted in 2018 identified 70 scientific articles providing supportive evidence that education guided by health behavioral theories was effective in promoting HPV vaccination [20]. However, only a few interventions on the perceptions of HPV vaccination have been published in China, and they involved limited sample sizes and did not involve health behavioral theories [21-23].

Currently, in China, there is a free HPV vaccination program in pilot regions among school girls less than 14 years old [24], but there are no effective strategies to improve HPV vaccine coverage among female college students who are at higher risk of HPV infection and have been suggested to be a "catch-up" population for HPV vaccination in many other countries. The information-motivation-behavioral skills (IMB) model is one of the commonly used theories on health promotion. To our knowledge, no previous study has been performed to evaluate IMB model–based online education for HPV vaccination among female college students from different regions in mainland China, with a follow-up duration of 3 months.

### Objectives

This study has the following 3 objectives: (1) to evaluate the feasibility and acceptability of the IMB model–based online intervention; (2) to examine whether and how this intervention improves HPV vaccination; and (3) to identify the barriers and facilitators of HPV vaccination among female college students in mainland China.

## *Methods*

### Study Design

The protocol for this study was published before the interventions were carried out [25]. Briefly, this study was a multicenter 2-arm cluster randomized trial. First-year female college students from 7 universities from different geographical locations in mainland China were recruited and randomly assigned 1:1 to either an IMB model–based intervention group or a waitlist control group based on their class. The intervention group received online education to promote the willingness to undergo HPV vaccination for 7 consecutive days, which was guided by the IMB model and conducted on a communication platform called DingTalk (Alibaba Group). The intervention group was compared with the waitlist control group in terms of the perceptions of HPV vaccination and willingness to undergo HPV vaccination. Recruitment of the study participants began in February 2020, and data collection was completed in July 2020. The whole process of participant recruitment, data collection, and intervention was conducted online, which coincided with the outbreak of COVID-19 in China, when Chinese college students were all self-isolated at home. The CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist was used to guide the reporting of our web-based survey [26]. The CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and

Mobile Health Applications and Online Telehealth) checklist is presented in Multimedia Appendix 1 [27].

## Recruitment and Eligibility

Participating students were recruited through notices on campus or advertisements in social media groups, posted in advance by partner teachers. Interested participants scanned the QR code on the notices or advertisements to fill in their class name and provide informed consent regarding the research objectives, requirements, procedures, benefits, and other study-related information. Potential participants were assessed for eligibility before the baseline survey. The eligibility criteria for the study were as follows: (1) female sex; (2) age ≥18 years; (3) first-year college student; (4) no vaccination contraindications; and (5) accessibility to computers or smartphones. Enrolled participants were given a link to either the intervention group or the control group that had already been set up in advance in DingTalk.

## Randomization and Blinding

In order to facilitate the management of participants and reduce loss of follow-up, enrolled participants were randomly assigned by class to the intervention and control groups. The investigator in each college, who was blinded to the identity of the participants, used computer software (Excel program, Microsoft Corp) to generate a series of random numbers for both the first-year Arts and Science major classes. Eligible participants were identified as either the intervention group or the control group based on the class name they entered when they scanned the QR code on the recruiting information. Participants, data analysts, and investigators were all blinded to the randomized allocation, and only the research assistants in each center were aware of the allocation.

## IMB Model–Based Intervention

### *Theoretical Framework*

The theoretical framework used to guide this online intervention was the IMB model [28]. This theory assumes that a person with rich knowledge will have the intention to practice healthy behaviors when he/she has motivation, the ability/skills to complete healthy behaviors, and self-efficacy, and the intention will be easily transformed into actual practice when objective conditions permit [19,29]. Therefore, we developed 2-day materials to popularize HPV knowledge, 2-day materials on situational stories to motivate participants to vaccinate themselves against HPV, and 3-day materials on objective skills with self-decision making, self-efficacy, and objective conditions for making an appointment and receiving the HPV vaccine. The above intervention materials were designed as online readable texts or videos that could be easily accessed by the target population.

### *Intervention Materials*

The intervention materials were developed by the research team based on the IMB model, and were uploaded and shared via the DingTalk platform by research assistants at each center. Upon randomization, the intervention group accessed the materials by scanning the QR code on each of the cover pages of the educational materials. During the daily intervention, we included quizzes on each day’s topics to check and consolidate the knowledge gained by the participants. It took about 10-15 minutes for the participants to read and learn the materials, depending on their learning ability. The forms, contents, and corresponding purposes of the IMB model–based education in this intervention are presented in Table 1.

**Table 1.** Cover pages, contents, and purposes of the information-motivation-behavioral skills model–based 7-day education.

| Serial number | Cover page | Content | Examples of the quizzes | Purpose |
|---|---|---|---|---|
| 1 |  | General facts about HPV[a], including HPV infection and related diseases. For example, there are more than a hundred different types of HPV, some of which are high risk and linked to the development of cancers, including cervical cancers. | 1. What is the transmission route of HPV?<br>2. True or false: HPV infection can cause condyloma acuminata, oropharyngeal cancer, cervical cancer, and anal cancer. | To provide a general overview of the topic, by attracting the subjects' attention with interesting animations. |
| 2 |  | Frequently asked questions and answers about HPV vaccine. For example, how does the HPV vaccine work? What are the recommended ages and populations for vaccination? | 1. True or false:<br>(1) HPV vaccine is a cervical cancer vaccine.<br>(2) The best time for HPV vaccination is before sex debut.<br>(3) Women who have been vaccinated against HPV do not need cervical screening. | To convey relevant messages to improve the subjects' understanding of HPV vaccines and vaccination. |
| 3 |  | A story of a female movie star who died of cervical cancer. Then, a case of a woman who missed HPV vaccination and cervical cancer screening, reached the terminal stage of cancer right after being found symptomatic, and died after painful treatment. Finally, the relevant facts about the screening, treatment, and prognosis of cervical cancer. | 1. True or false: Persistent HPV infection may cause cervical cancer.<br>2. What is the recommended age for cervical cancer screening? | To use real-life experiences to stimulate the participants' fear and susceptibility to cervical cancer and arouse their desire to be vaccinated against HPV. |
| 4 |  | Risk factors and early symptoms of cervical cancer, and ways to prevent and control cervical cancer. | 1. Regarding the high-risk factors of cervical cancer, which of the following is wrong?<br>A. Premature birth and fertility<br>B. Premature sex<br>C. Hormone replacement therapy<br>D. Sexual disorder | To call for HPV vaccination and regular cervical cancer screening, in order to arouse the desire of participants to get vaccinated against HPV. |
| 5 |  | The concept and practice of decision-making and effective communication. | 1. Application information: Please select the views that support HPV vaccination (multiple choices)<br>A. I am at risk of contracting HPV.<br>B. I am afraid of needle tingling.<br>C. HPV infection will affect my daily life.<br>D. The price of vaccines is too expensive. | To make firm the participants' determination to receive HPV vaccines, and communicate effectively with parents and friends to obtain support. |
| 6 |  | The concept and function of self-efficacy, evaluation of self-efficacy, and ways to improve self-efficacy. | What are the ways to improve self-efficacy (multiple choices)?<br>1. Try to stick to a good habit.<br>2. Concentrate on what you do.<br>3. Evaluate the pros and cons and make a decision.<br>4. Self-discipline and self-motivation. | To provide guidance on how to improve self-efficacy and turn the idea of HPV vaccination into action. |
| 7 |  | Provide participants with the objective skills needed for HPV vaccination in the form of invitation letters designed with HTML5 front-end technology, such as appointment platform, price, vaccination venue, etc. | The last page is linked to the online questionnaire after the intervention. | To provide detailed objective information required for HPV vaccination, making HPV vaccination more accessible and convenient. |

[a] HPV: human papillomavirus.

## Waitlist Control Group

The study was carried out when COVID-19 broke out worldwide at the beginning of 2020. In order to increase the compliance of the control group and reduce the probability of compromising the blind nature of the study, the control group was given 7 days of information on COVID-19 prevention, which was organized and presented in the same format and platform as the educational materials in the intervention group. At the end of the study, the waitlist control group received the same educational materials as the intervention group.

### Outcomes and Measurements

The internet-based questionnaire surveys were administered at 4 time points (baseline, immediately after the intervention, and 1 month and 3 months after the intervention), and participants were given notebooks and pens with the logo of the research institution as incentives.

Background characteristics, including age, major in college, ethnicity, residence, parental residence, education and marital status, monthly living expenses (RMB), family/friends with any cancer, ever received sexual education, currently in a romantic relationship, sexual debut, and attitude toward premarital sex, were measured in this study. These variables are necessary information, basic information, or information related to HPV vaccination as shown in previous literature [30].

The primary outcome measures were self-reported willingness to undergo HPV vaccination and uptake of HPV vaccination. At baseline and each follow-up, the participants were asked "Are you willing to get the HPV vaccine in the future?" and "Have you been vaccinated against HPV?" with "Yes" and "No" response options. It should be noted that there were some differences in the way we asked questions regarding the willingness to undergo HPV vaccination. The 3-month questioning time frame was in the "future," while the baseline and other 2 follow-up questioning time frames were "within the study period." The differences in outcomes are detailed in the Results section.

The secondary outcome measures were the information/knowledge, motivation, and behavioral skills regarding HPV vaccination, which were designed based on the IMB model. Among them, the information part consisted of 11 questions. For example, "HPV is related to the development of cervical cancer," with answer options "Agree," "Disagree," and "Do not know" (Cronbach α=.78) [30-33]. Motivation for vaccination was measured by 19 questions. For example, "Getting vaccinated for HPV will help protect me from HPV infection" (Cronbach α=.71) [31,34-36]. Behavioral skills were measured by 10 questions. For example, "I feel confident in my ability to get vaccinated for HPV, even if it is expensive" (Cronbach α=.88) [33,35]. The answers for these items were measured on a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neither disagree nor agree, 4=agree, and 5=strongly agree). A description of the baseline results for this study has been published [37].

In the 3rd month after the intervention, the participants were asked about their perceptions of the barriers and facilitators of HPV vaccination, such as the reasons for not receiving the HPV vaccine; choices of the HPV vaccine; opinions on promoting HPV vaccination in China; and willingness to receive the HPV vaccine under different scenarios.

### Ethics Approval

This study was approved by the Institutional Review Board of the Chinese Center for Disease Control and Prevention on October 24, 2019 (approval number: 201918-01).

### Statistical Analysis

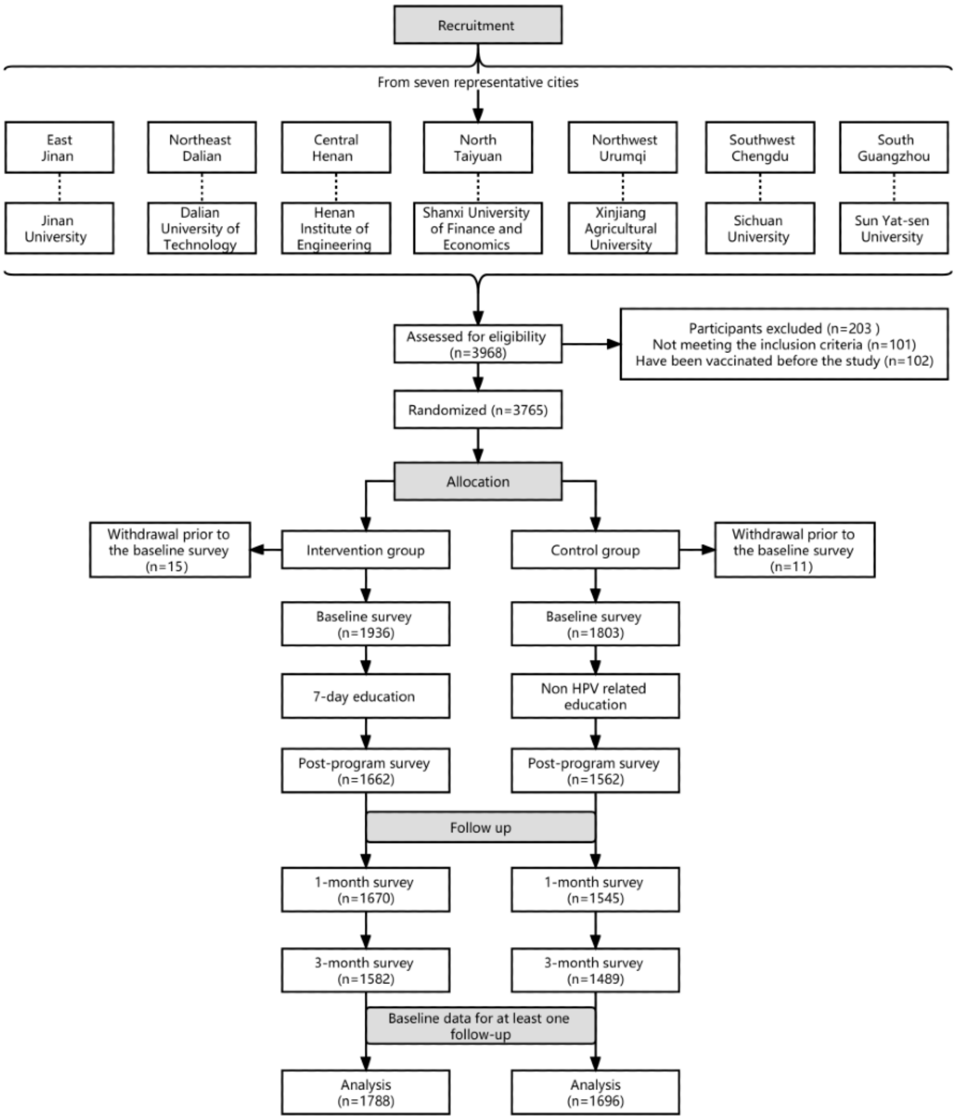
IBM SPSS Statistics 23.0 (IBM Corp) was used to process the data and conduct statistical comparisons between the intervention and control groups. Independent *t* tests and chi-square tests were used to compare the distributions of the continuous and categorical variables, respectively. Analyses were conducted based on an intention-to-treat approach, and statistical significance was set at $P<.05$ (2-sided). The effects of the IMB model–based intervention on knowledge, motivation, behavioral skills, and willingness regarding HPV vaccination were examined using generalized estimating equations (GEEs). We included significant variables in baseline chi-square analysis, and assessed group (intervention and control), time (baseline, immediately after the intervention, and 1 month and 3 months after the intervention), and time × group interaction, with the time × group interaction indicating a differential change by group from baseline to the end of the trial.

## Results

### Overview

The flowchart for participant recruitment is presented in Figure 1. From February 2020 to March 2020, a total of 4051 female college students were recruited and screened, among whom 83 refused to participate in the study, 101 did not meet the inclusion criteria, and 102 reported that they had been vaccinated with the HPV vaccine prior to the study. Among the participants who met the inclusion criteria, the HPV vaccination rate was 2.6% (102/3867). A total of 3765 participants signed the informed consent and completed the baseline questionnaire (T-baseline). However, 26 participants were not grouped and withdrew from the study. There were eventually 1936 participants in the intervention group and 1803 participants in the control group. A total of 3224 (86.2%) participants completed the postintervention questionnaire (T-postintervention). In the intervention group, 1662 participants completed the IMB model–based intervention, and 1562 participants in the control group received non-HPV–related content in the same period. In addition, a total of 3215 (86.0%) participants (1670 in the intervention group and 1545 in the control group) completed the follow-up assessment 1 month after the intervention (T-1 month). Finally, a total of 3071 (82.1%) participants completed the 3-month follow-up evaluation (T-3 months), including 1582 and 1489 participants in the intervention and control groups, respectively. The proportions of participants in the intervention group and control group who completed the T-postintervention assessment were 85.9% (1662/1936) and 86.6% (1562/1803), respectively, and the difference was not statistically significant ($P=.49$). The completion rates of the T-1 month assessment in the intervention and control groups were 86.2% (1670/1936) and 86.6% (1562/1803), respectively, and the difference was not statistically significant ($P=.62$). The completion rates of the T-3 months assessment in the intervention and control groups were 81.7% (1582/1936) and 82.6% (1489/1803), respectively, with no statistically significant difference (*P*=.49).

**Figure 1.** The study flowchart.



## Participant Characteristics

The participants with at least one follow-up record were included in our analysis. A total of 3484 respondents were identified, and the mean age was 19.12 (SD 0.73) years. Over half of the respondents (1848/3484, 53.0%) were majoring in Arts in college, and 87.7% (3057/3484) were Han Chinese. Most of the respondents (2207/3484, 63.4%) and their parents (2189/3484, 62.8%) lived in urban areas during the study period. Regarding the educational background of the parents, less than half (1591/3484, 45.7%) of them had a junior high school or lower education, and only 24.6% (856/3484) of the parents had a college degree or above. Most of the respondents' parents (2819/3484, 80.9%) were married at the time of the research, while 6.9% (239/3484) were divorced, 9.8% (340/3484) were unmarried cohabiting, and 2.5% (86/3484) were widowed. Overall, 71.3% (2483/3484) had monthly living expenses ranging from 1000 to 2000 yuan (US$ 144.35 to 288.71), and about a third (842/3484, 24.2%) of all respondents had relatives or friends with cancer. Moreover, 79.2% (2758/3484) of the respondents self-reported having ever received sexual education, while 16.9% (590/3484) were currently in a romantic relationship. Only 2.8% (97/3484) had sexual debut, while 37.6% (1310/3484) said they could accept premarital sex. As shown in Table 2, both groups were comparable in terms of their sociodemographic characteristics, except previous sexual intercourse, for which there were more participants in the intervention group than in the control group.

**Table 2.** Baseline characteristics of the participants.

| Variable | All (N=3484) | Intervention group (N=1788) | Control group (N=1696) | $\chi^2$ (df) | *P* value |
|---|---|---|---|---|---|
| Age (years), mean (SD) | 19.12 (0.73) | 19.12 (0.72) | 19.11 (0.73) | 0.34[a] (3482) | .73 |
| **Major in college, n (%)** | | | | 0.73 (1) | .39 |
| Arts | 1848 (53.0) | 961 (53.8) | 887 (52.3) | | |
| Science | 1636 (47.0) | 827 (46.3) | 809 (47.7) | | |
| **Ethnicity, n (%)** | | | | 0.25 (1) | .62 |
| Han | 3057 (87.7) | 1564 (87.5) | 1493 (88.0) | | |
| Other | 427 (12.3) | 224 (12.5) | 203 (12.0) | | |
| **Location, n (%)** | | | | 1.32 (1) | .25 |
| Urban | 2207 (63.4) | 1149 (64.3) | 1058 (62.4) | | |
| Rural | 1277 (36.7) | 639 (35.7) | 638 (37.6) | | |
| **Parental location, n (%)** | | | | 1.05 (1) | .31 |
| Urban | 2189 (62.8) | 1138 (63.7) | 1051 (62.0) | | |
| Rural | 1295 (37.2) | 650 (36.4) | 645 (38.0) | | |
| **Parental education, n (%)** | | | | 0.97 (2) | .62 |
| Junior high school or less | 1591 (45.7) | 1428 (79.9) | 1391 (82.0) | | |
| Senior high school or technical secondary school | 1037 (29.8) | 120 (6.7) | 119 (7.0) | | |
| College or more | 856 (24.6) | 189 (10.6) | 151 (8.9) | | |
| **Marital status of parents, n (%)** | | | | 5.29 (3) | .15 |
| Married | 2819 (80.9) | 51 (2.9) | 35 (2.1) | | |
| Divorced/separated | 239 (6.9) | 831 (46.5) | 760 (44.8) | | |
| Unmarried cohabiting | 340 (9.8) | 524 (29.3) | 513 (30.3) | | |
| Widowed | 86 (2.5) | 433 (24.2) | 423 (24.9) | | |
| **Monthly living expenses (RMB)[b], n (%)** | | | | 5.49 (2) | .06 |
| <1000 | 720 (20.7) | 392 (21.9) | 328 (19.3) | | |
| 1000-2000 | 2483 (71.3) | 1243 (69.5) | 1240 (73.1) | | |
| >2000 | 281 (8.1) | 153 (8.6) | 128 (7.6) | | |
| **Family/friends with any cancer, n (%)** | | | | 0.78 (1) | .38 |
| Yes | 842 (24.2) | 421 (23.6) | 421 (24.8) | | |
| No | 2642 (75.8) | 1367 (76.5) | 1275 (75.2) | | |
| **Ever received sexual education, n (%)** | | | | 0.40 (1) | .53 |
| Yes | 2758 (79.2) | 1423 (79.6) | 1335 (78.7) | | |
| No | 726 (20.8) | 365 (20.4) | 361 (21.3) | | |
| **Currently in a romantic relationship, n (%)** | | | | 0.38 (1) | .54 |
| Yes | 590 (16.9) | 296 (16.6) | 294 (17.3) | | |
| No | 2894 (83.1) | 1492 (83.5) | 1402 (82.7) | | |
| **Had sexual debut, n (%)** | | | | 4.43 (1) | .04[c] |
| Yes | 97 (2.8) | 60 (3.4) | 37 (2.2) | | |
| No | 3387 (97.2) | 1728 (96.6) | 1659 (97.8) | | |
| **Attitude toward premarital sex, n (%)** | | | | 0.29 (1) | .59 |
| Yes | 1310 (37.6) | 680 (38.0) | 630 (37.2) | | |

| Variable | All (N=3484) | Intervention group (N=1788) | Control group (N=1696) | $\chi^2$ (df) | *P* value |
|---|---|---|---|---|---|
| No | 2174 (62.4) | 1108 (62.0) | 1066 (62.9) | | |

[a] Analysis for *t* test.

[b] A currency exchange rate of 1 RMB=US $0.14435 is applicable.

[c] Significant *P*<.05.

## Evaluation of the Intervention

Baseline differences were discovered for previous sexual intercourse alone, and as such, all subsequent analyses were adjusted for previous sexual experience (Table 3).

**Table 3.** Generalized estimating equation model of willingness and the 3 dimensions of the information-motivation-behavioral skills model regarding human papillomavirus vaccination between the intervention and control groups.

| Variable[a] | Willingness to receive the HPV[b] vaccine | | Information | | Motivation | | Behavioral skills | |
|---|---|---|---|---|---|---|---|---|
| | β | *P* value | β | *P* value | β | *P* value | β | *P* value |
| **Sexual debut** | | | | | | | | |
| No | Reference | | Reference | | Reference | | Reference | |
| Yes | 1.025 | <.001[c] | 1.037 | <.001[c] | 0.171 | <.001[c] | 0.137 | .01[c] |
| **Time[d]** | | | | | | | | |
| T-baseline | Reference | | Reference | | Reference | | Reference | |
| T-postintervention | −0.019 | .67 | 0.391 | <.001[c] | −0.055 | <.001[c] | −0.029 | .006[c] |
| T-1 month | −0.055 | .27 | 0.628 | <.001[c] | −0.065 | <.001[c] | −0.043 | <.001[c] |
| T-3 months | 1.476 | <.001[c] | 0.762 | <.001[c] | −0.035 | <.001[c] | −0.017 | .23 |
| **Group** | | | | | | | | |
| Control | Reference | | Reference | | Reference | | Reference | |
| Intervention | 0.042 | .56 | 0.103 | .26 | 0.002 | .84 | 0.011 | .52 |
| **Time[d] × group** | | | | | | | | |
| T-baseline × intervention/control | Reference | | Reference | | Reference | | Reference | |
| T-postintervention × intervention/control | 0.320 | <.001[c] | 2.202 | <.001[c] | 0.115 | <.001[c] | 0.104 | <.001[c] |
| T-1 month × intervention/control | 0.183 | .01[c] | 1.892 | <.001[c] | 0.085 | <.001[c] | 0.062 | <.001[c] |
| T-3 months × intervention/control | 0.140 | .13 | 1.688 | <.001[c] | 0.053 | <.001[c] | 0.059 | .003[c] |

[a] The generalized estimating equation model was adjusted with the significant variable in the chi-square analysis of baseline data, namely “sexual debut.”

[b] HPV: human papillomavirus.

[c] Significant *P*<.05.

[d] Time points: baseline (T-baseline), immediately after the intervention (T-postintervention), 1 month after the intervention (T-1 month), and 3 months after the intervention (T-3 months).

### *Willingness to Receive the HPV Vaccine*

The GEE revealed the simple effects of time (T-3 months vs T-baseline, β=1.476; *P*<.001) and a significant time × group interaction for T-postintervention vs T-baseline (β=0.320; *P*<.001) and T-1 month vs T-baseline (β=0.183; *P*=.011), but no group effect (intervention group vs control group; β=0.042, *P*=.56). Compared with the control group, the intervention group showed a significant increase in the willingness to undergo HPV vaccination immediately and 1 month after the intervention.

### *HPV Information*

The GEE revealed the simple effects of time (T-postintervention, T-1 month, and T-3 months vs T-baseline; β=0.391, 0.628, and 0.762, respectively; *P*<.001) and a significant time × group interaction for T-postintervention, T-1 month, and T-3 months vs T-baseline (β=2.202, 1.892, and 1.688, respectively; *P*<.001), but no group effect (intervention group vs control group; β=0.103; *P*=.26). Compared with the control group, the intervention group showed a significant increase in HPV information after the intervention.

### *Motivation for HPV Vaccination*

The GEE revealed the simple effects of time (T-postintervention, T-1 month, and T-3 months vs T-baseline; β=−0.055, −0.065, and −0.035, respectively; *P*<.001) and a significant time × group interaction for T-postintervention, T-1 month, and T-3 months vs T-baseline (β=0.115, 0.085, and 0.053, respectively; *P*<.001), but no group effect (intervention group vs control group; β=0.002; *P*=.84). Compared with the control group, the intervention group showed a significant increase in motivation for HPV vaccination after the intervention.

### *Behavioral Skills for HPV Vaccination*

The GEE revealed the simple effects of time for T-postintervention vs T-baseline (β=−0.029; *P*=.006) and for T-1 month vs T-baseline (β=−0.043; *P*<.001). A significant time × group interaction was found for T-postintervention and T-1 month vs T-baseline (β=0.104 and 0.062, respectively; *P*<.001) and for T-3 months vs T-baseline (β=0.059; *P*=.003), but there was no group effect (intervention group vs control group; β=0.011; *P*=.52). Compared with the control group, the intervention group showed a significant increase in behavioral skills for HPV vaccination after the intervention.

## Intervention Effects on the Willingness to Undergo HPV Vaccination and Practice of HPV Vaccination

Table 4 shows the main effect of the intervention on the willingness to undergo HPV vaccination from baseline to follow-up. At baseline, the willingness rates to undergo HPV vaccination in the intervention and control groups were 33.33% and 31.96%, respectively, and there was no significant difference between the 2 groups ($\chi^2_1$=0.75; *P*=.39). Immediately after the intervention, the willingness rates to undergo HPV vaccination in the intervention and control groups were 40.39% and 31.56%, respectively, with statistically significant differences between the 2 groups and within the intervention group from T-baseline to T-postintervention ($\chi^2_1$=27.11; *P*<.001; and $\chi^2_1$=15.43; *P*<.001, respectively). At T-1 month, the willingness rates to receive the HPV vaccine in the intervention and control groups were 35.65% and 31.34%, respectively. At this point, the difference between the 2 groups was statistically significant ($\chi^2_1$=6.64; *P*=.01), but there was no significant difference within the intervention group from T-baseline to T-1 month ($\chi^2_1$=1.40; *P*=.24). At T-3 months, the willingness rates to undergo HPV vaccination in the intervention and control groups were 70.59% and 66.17%, respectively, with statistically significant differences between the 2 groups and within the intervention group from T-baseline to T-3 months ($\chi^2_1$=6.81; *P*=.01; and $\chi^2_1$=368.59; *P*<.001, respectively).

At different time points, the HPV vaccination rates in the intervention and control groups were as follows: T-postintervention, 0.48% and 0.19%, respectively ($\chi^2_1$=1.98; *P*=.16); T-1 month, 0.72% and 0.65%, respectively ($\chi^2_1$=0.06; *P*=.81); T-3 months, 2.21% and 1.75%, respectively ($\chi^2_1$=0.86; *P*=.36).

**Table 4.** Main effect of the intervention on the willingness to undergo human papillomavirus vaccination from baseline to follow-up.

| Variable and time[a] | Intervention group | | $\chi^{2}$[b] (df) | *P* value[b] | Control group | | $\chi^{2}$[c] (df) | *P* value[c] |
|---|---|---|---|---|---|---|---|---|
| | Yes, n/N (%) | No, n/N (%) | | | Yes, n/N (%) | No, n/N (%) | | |
| **Willingness to receive the HPV[d] vaccine** | | | | | | | | |
| T-baseline[e] | 596/1788 (33.3) | 1192/1788 (66.7) | N/A[f] | N/A | 542/1696 (32.0) | 1154/1696 (68.0) | 0.75 (1) | .39 |
| T-postintervention[e] | 668/1654 (40.4) | 986/1654 (59.6) | 15.43 (1) | <.001[g] | 492/1559 (31.6) | 1067/1559 (68.4) | 27.11 (1) | <.001[g] |
| T-1 month[e] | 591/1658 (35.7) | 1067/1658 (64.3) | 1.40 (1) | .24 | 481/1535 (31.3) | 1054/1535 (68.7) | 6.64 (1) | .01[g] |
| T-3 months[h] | 1092/1547 (70.6) | 455/1547 (29.4) | 368.59 (1) | <.001[g] | 968/1463 (66.2) | 495/1463 (33.8) | 6.81 (1) | .01[g] |

[a] Time points: baseline (T-baseline), immediately after the intervention (T-postintervention), 1 month after the intervention (T-1 month), and 3 months after the intervention (T-3 months).

[b] Comparison between baseline and different follow-up time points in the intervention group.

[c] Comparison between the intervention and control groups at baseline and each follow-up time point.

[d] HPV: human papillomavirus.

[e] Are you willing to get the HPV vaccine in recent months?

[f] N/A: not applicable.

[g] Significant *P*<.05.

[h] Are you willing to get the HPV vaccine in the future?

## Intervention Effects on IMB Model Variables

### HPV-Related Information

Compared with the control group, the participants in the intervention group demonstrated a significant improvement in knowledge scores from baseline to any posttest time point (*P*<.001). In addition, there were significant changes in knowledge scores from baseline to any posttest time point (*P*<.001) within the intervention group (Figure 2A; Table 5).

**Figure 2.** Trends of the mean scores of information (A), motivation (B), and behavioral skills (C) regarding human papillomavirus (HPV) vaccination in the intervention and control groups over time.

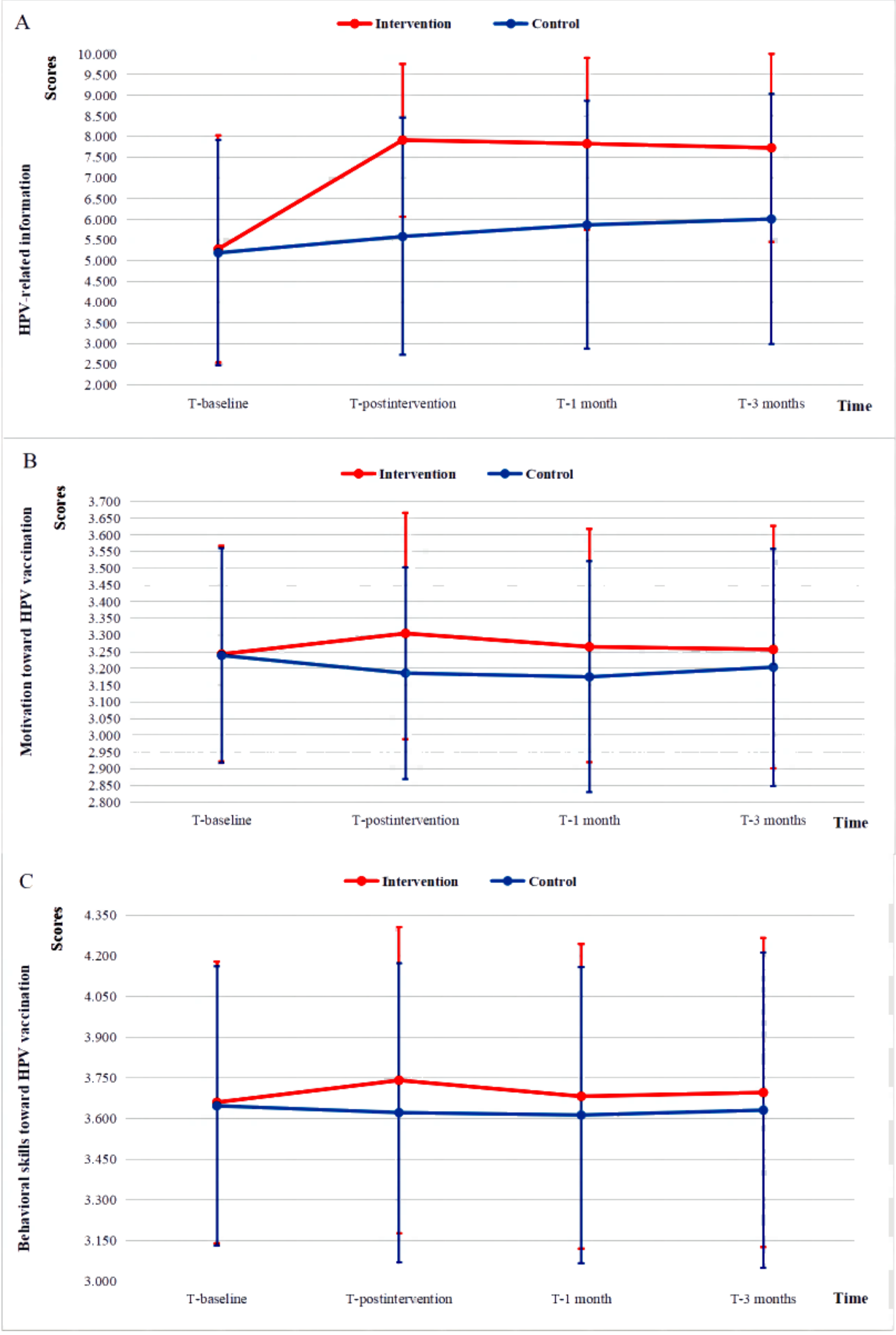


**Table 5.** Main effect of the intervention on the mean scores of the information-motivation-behavioral skills model constructs regarding human papillomavirus vaccination from baseline to follow-up.

| IMB[a] model constructs and time[b] | Intervention group, mean score (SD) | Mean difference within the group (95% CI) | *t*[c] (df) | *P* value[c] | Control group, mean score (SD) | Mean difference between groups (95% CI) | *t*[d] (df) | *P* value[d] |
|---|---|---|---|---|---|---|---|---|
| **Information** | | | | | | | | |
| T-baseline | 5.29 (2.74) | N/A[e] | N/A | N/A | 5.20 (2.72) | 0.93 (−0.09 to 0.27) | 1.00 (3482) | .32 |
| T-postintervention | 7.92 (1.84) | 2.58 (2.43 to 2.74) | 32.70 (3448) | <.001[f] | 5.59 (2.86) | 2.33 (2.16 to 2.49) | 27.29 (3222) | <.001[f] |
| T-1 month | 7.83 (2.07) | 2.51 (2.34 to 2.67) | 30.52 (3456) | <.001[f] | 5.87 (2.99) | 1.96 (1.78 to 2.14) | 21.46 (3213) | <.001[f] |
| T-3 months | 7.73 (2.27) | 2.40 (2.22 to 2.57) | 27.25 (3368) | <.001[f] | 6.01 (3.02) | 1.72 (1.53 to 1.91) | 17.76 (3069) | <.001[f] |
| **Motivation** | | | | | | | | |
| T-baseline | 3.24 (0.32) | N/A | N/A | N/A | 3.24 (0.32) | 0.00 (−0.02 to 0.03) | 0.39 (3482) | .70 |
| T-postintervention | 3.31 (0.36) | 0.06 (0.03 to 0.08) | 4.81 (3448) | <.001[f] | 3.19 (0.32) | 0.12 (0.19 to 0.14) | 10.00 (3222) | <.001[f] |
| T-1 month | 3.27 (0.35) | 0.02 (−0.01 to 0.04) | 1.50 (3456) | .13 | 3.18 (0.35) | 0.09 (0.07 to 0.11) | 7.31 (3213) | <.001[f] |
| T-3 months | 3.26 (0.37) | 0.01 (−0.02 to 0.03) | 0.59 (3368) | .56 | 3.20 (0.36) | 0.05 (0.03 to 0.08) | 4.01 (3069) | <.001[f] |
| **Behavioral skills** | | | | | | | | |
| T-baseline | 3.66 (0.52) | N/A | N/A | N/A | 3.65 (0.52) | 0.01 (−0.02 to 0.05) | 0.74 (3482) | .46 |
| T-postintervention | 3.74 (0.57) | 0.07 (0.03 to 0.10) | 3.70 (3448) | <.001[f] | 3.62 (0.55) | 0.12 (0.08 to 0.16) | 6.04 (3222) | <.001[f] |
| T-1 month | 3.68 (0.56) | 0.01 (−0.03 to 0.05) | 0.60 (3456) | .55 | 3.61 (0.55) | 0.067 (0.03 to 0.11) | 3.54 (3213) | <.001[f] |
| T-3 months | 3.70 (0.57) | 0.02 (−0.02 to 0.05) | 0.85 (3368) | .39 | 3.63 (0.58) | 0.07 (0.02 to 0.11) | 3.16 (3069) | <.001[f] |

[a] IMB: information-motivation-behavioral skills.

[b] Time points: baseline (T-baseline), immediately after the intervention (T-postintervention), 1 month after the intervention (T-1 month), and 3 months after the intervention (T-3 months).

[c] Comparison between baseline and different follow-up time points in the intervention group.

[d] Comparison between the intervention and control groups at baseline and each follow-up time point.

[e] N/A: not applicable.

[f] Significant *P*<.05.

### *Motivation for HPV Vaccination*

Compared with the control group, the participants in the intervention group demonstrated a significant improvement in motivation from baseline to any posttest time point (*P*<.001). However, in the intervention group, there were no significant changes in motivation scores from T-baseline to T-1 month or T-3 months (*P*=.13 and *P*=.56, respectively), except for a significant change from T-baseline to T-postintervention (*P*<.001) (Figure 2B; Table 5).

### *Behavioral Skills for HPV Vaccination*

Similar to the trend for motivation, compared with the control group, the participants in the intervention group demonstrated a significant improvement in behavioral skills from baseline to any posttest time point (*P*<.001). However, in the intervention group, there were no significant changes in scores of behavioral skills from T-baseline to T-1 month or T-3 months (*P*=.55 and *P*=.39, respectively), except for a significant change from T-baseline to T-postintervention (*P*<.001) (Figure 2C; Table 5).

## Perceptions of the Barriers and Facilitators of Receiving the HPV Vaccine

Table 6 shows that 2060 (68.4%) participants were willing to undergo HPV vaccination and 2788 (90.52%) participants would recommend HPV vaccination to their relatives and friends. The reasons for not undergoing or recommending HPV vaccination included “worried about its side effects” (471/1694, 27.8%), “expensive” (433/1694, 25.6%), “unsure of its safety and effectiveness” (406/1694, 24.0%), “lack of knowledge about the HPV vaccine” (318/1694, 18.8%), “have no sexual behavior” (279/1694, 16.5%), and “possible needle injury” (178/1694, 10.5%).

**Table 6.** Perceptions of the barriers and facilitators of receiving the human papillomavirus vaccine.

| Variable | All, n (%) | Intervention group, n (%) | Control group, n (%) |
|---|---|---|---|
| **Reasons for not receiving the HPV[a] vaccine** | | | |
| **Willing to undergo HPV vaccination (N=3010; I[b]: 1547, C[c]: 1463)** | | | |
| Yes | 2060 (68.4) | 1092 (70.6) | 968 (66.2) |
| No | 950 (31.6) | 455 (29.4) | 495 (33.8) |
| **Willing to recommend HPV vaccination (N=3071; I: 1582, C: 1489)** | | | |
| Yes | 2780 (90.5) | 1447 (91.5) | 1333 (89.5) |
| No | 291 (9.5) | 135 (8.5) | 156 (10.5) |
| **Reasons for not undergoing or recommending HPV vaccination (N=1694; I: 863, C: 831)** | | | |
| Do not know much about the HPV vaccine | 318 (18.8) | 119 (13.8) | 199 (24.0) |
| Unsure of its safety and effectiveness | 406 (24.0) | 171 (19.8) | 235 (28.3) |
| Worried about its side effects | 471 (27.8) | 210 (24.3) | 261 (31.4) |
| Possible needle injury | 178 (10.5) | 82 (9.5) | 96 (11.6) |
| Expensive | 433 (25.6) | 228 (26.4) | 205 (24.7) |
| Too many needles | 123 (7.3) | 68 (7.9) | 55 (6.6) |
| No sexual behavior | 279 (16.5) | 142 (16.5) | 137 (16.5) |
| Others | 19 (1.1) | 14 (1.6) | 5 (0.6) |
| **Opinions on promoting HPV vaccination (N=3071; I: 1582, C: 1489)** | | | |
| **The promotion of the HPV vaccine will be well accepted** | | | |
| Yes | 2760 (89.9) | 1425 (90.1) | 1335 (89.7) |
| No | 311 (10.1) | 157 (9.9) | 154 (10.3) |
| **The biggest concern about the HPV vaccine** | | | |
| Efficacy/effectiveness | 507 (16.5) | 259 (16.4) | 248 (16.7) |
| Safety | 856 (27.9) | 436 (27.6) | 420 (28.2) |
| Side effects | 543 (17.7) | 254 (16.1) | 289 (19.4) |
| Price | 786 (25.6) | 425 (26.9) | 361 (24.2) |
| Procedure for appointment and HPV vaccination | 250 (8.1) | 135 (8.5) | 115 (7.7) |
| Nothing to worry about | 94 (3.1) | 57 (3.6) | 37 (2.5) |
| **The barriers for promoting HPV vaccination** | | | |
| Insufficient message from the government | 1660 (54.1) | 821 (51.9) | 839 (56.4) |
| Too expensive | 2330 (75.9) | 1230 (77.8) | 1100 (73.9) |
| Not covered by medical insurance | 1718 (55.9) | 874 (55.3) | 844 (56.7) |
| Not included in national immunization programs | 1587 (51.7) | 806 (51.0) | 781 (52.5) |
| Inconvenient to schedule vaccination | 1218 (39.7) | 611 (38.6) | 607 (40.8) |
| Insufficient experience of doctors | 770 (25.1) | 360 (22.8) | 410 (27.5) |
| Vaccine hesitancy | 769 (25.0) | 374 (23.6) | 395 (26.5) |
| HPV vaccine–related stigma | 550 (17.9) | 266 (16.8) | 284 (19.1) |
| **How to promote HPV vaccination** | | | |
| Improve doctor’s recommendation | 2341 (76.2) | 1189 (75.2) | 1152 (77.4) |
| Media publicity | 1793 (58.4) | 908 (57.4) | 885 (59.4) |
| Make it convenient to schedule vaccination | 1994 (64.9) | 1036 (65.5) | 958 (64.3) |
| Provide the vaccine at a reasonable price | 2453 (79.9) | 1271 (80.3) | 1182 (79.4) |

| Variable | All, n (%) | Intervention group, n (%) | Control group, n (%) |
|---|---|---|---|
| **The most preferable type of HPV vaccine (N=3010; I: 1547, C: 1463)** | | | |
| 2-valent HPV (Cervarix, GlaxoSmithKline ) | 552 (18.3) | 287 (18.6) | 265 (18.1) |
| 4-valent HPV (Gardasil, Merck and Co, Inc) | 425 (14.1) | 224 (14.5) | 201 (13.7) |
| 9-valent HPV (Gardasil 9, Merck and Co, Inc) | 947 (31.5) | 489 (31.6) | 458 (31.3) |
| 2-valent HPV (Cecolin, Innovax) | 1087 (36.1) | 548 (35.4) | 539 (36.8) |
| **Willingness to receive the HPV vaccine in the future under different scenarios (N=3010; I: 1547, C: 1463)** | | | |
| **If the price is affordable** | | | |
| Yes | 2839 (94.3) | 1465 (94.7) | 1374 (93.9) |
| No | 171 (5.7) | 82 (5.3) | 89 (6.1) |
| **If it is covered by medical insurance** | | | |
| Yes | 2862 (95.1) | 1471 (95.1) | 1391 (95.1) |
| No | 148 (4.9) | 76 (4.9) | 72 (4.9) |
| **If it is included in national immunization programs** | | | |
| Yes | 2887 (95.9) | 1492 (96.4) | 1395 (95.4) |
| No | 123 (4.1) | 55 (3.6) | 68 (4.6) |

[a] HPV: human papillomavirus.

[b] I: number of participants in the intervention group.

[c] C: number of participants in the control group.

Among the 3071 complaint participants, most of them (2760/3071, 89.9%) believed that the promotion of the HPV vaccine would be well accepted. The major concerns about the HPV vaccine were "safety" (856/3071, 27.9%) and "price" (786/3071, 25.6%), followed by "side effects" (543/3071, 17.7%) and "efficacy" (507/3071, 16.5%). In addition, there were 250 (8.1%) participants concerned about the "procedure for appointment and uptake of the HPV vaccine," and only 94 (3.1%) participants did not express any concerns. Regarding the barriers for promoting HPV vaccination, "too expensive" (2330/3071, 75.9%), "not being covered by medical insurance" (1718/3071, 55.9%), "insufficient message from the government" (1660/3071, 54.1%), and "not being included in national immunization programs" (1587/3071, 51.7%) were major concerns. Each of the measures to promote HPV vaccination in China received more than 50% agreement from the participants, including "reasonable price" (2453/3071, 79.9%), "doctors' recommendation" (2341/3071, 76.2%), "be convenient to schedule vaccination" (1994/3071, 64.9%), and "media publicity" (1793/3071, 58.4%).

Interestingly, the most preferable type of HPV vaccine we found was the domestically produced 2-valent HPV vaccine (1087/3010, 36.1%), followed by the 9-valent HPV vaccine (947/3010, 31.5%), the 2-valent HPV vaccine (552/3010, 18.3%), and the 4-valent HPV vaccine (425/3010, 14.1%), which were all produced abroad.

The willingness rate to receive the HPV vaccine was 94.3% (2839/3010) if the price was affordable. In addition, it increased to 95.1% (2862/3010) if the HPV vaccination was covered by medical insurance, and it increased to 95.9% (2887/3010) if the HPV vaccination was included in national immunization programs, compared with 68.4% (2060/3010) unconditional willingness.

# Discussion

## Principal Findings

### *Overview*

This study investigated the effectiveness of an IMB model–based online education program for HPV vaccination among female students from 7 universities in mainland China. To our knowledge, this is the first study in China to evaluate the effect of an online health education program guided by a health behavioral theory on improving the awareness, attitudes, and willingness regarding HPV vaccination with a randomized intervention and multiple follow-ups.

### *Changes in the Willingness to Undergo HPV Vaccination*

After the education program, there was a significant increase in the willingness to be vaccinated against HPV from baseline in the intervention group. Moreover, the willingness to undergo HPV vaccination was significantly higher in the intervention group than in the control group at the end of the study. In line with the findings of previous studies [21,38,39], the effect of the health education program peaked right after the intervention and then gradually diminished over time, but finally, the residual effect of the health education program could still be observed. In this study, participants' motivation toward HPV vaccination was obviously activated right after the tailored IMB model–based education program, which was consistent with the finding of a previous similar study among impulsive youth [40]. As the passion fades, the participants might begin to think rationally and objectively based on what they had learned about

HPV-related issues. They might search for information to expand their knowledge of the HPV vaccine and then try to make an appointment for vaccination. During this process, they are likely to be concerned with the side effects and relatively high price of the HPV vaccine, and might find it difficult to make an appointment for HPV vaccination, which would curb their willingness to receive the vaccine. This study was conducted during the early stage of the COVID-19 pandemic. Due to the constraints of home isolation and traffic control, and the fact that medical resources were tilted toward the prevention and treatment of COVID-19 infection, it was more difficult for the participants to make an appointment and receive the HPV vaccine than usual [41,42]. Therefore, the willingness to be vaccinated against HPV decreased to the baseline level 1 month after the education program. Finally, at the end of the study, participants in the intervention and control groups self-reported a higher level of willingness to be vaccinated against HPV than that at baseline, and the willingness was significantly higher in the intervention group than in the control group. This reminds us that even if education cannot achieve the expected effect in the short term, the seeds of health education will take root and sprout in the participants' brains, which will trigger thinking in some people and may guide their decision-making in the future. Additionally, these results imply that re-education is necessary to sustain the educational effect.

### Intervention Effects on IMB Variables

The daily education and matching quizzes in the intervention period, which contained questions in the information part of the questionnaire, probably resulted in a deeper understanding of HPV-related issues in the intervention group compared with the control group, even at 3 months after the education program. In addition, the results also indicated that this study was readily acceptable to the participants in terms of both educational contents and forms, and this coincided with the feedback comments and suggestions from the participants at the end of the study (data are not presented in this manuscript).

Different from the trend of knowledge over time, motivation and behavioral skills regarding HPV vaccination only showed significant differences in the intervention group immediately after the education program. In this study, a 2-day education program addressing the role of preventing HPV infection and negative treatment experiences was given to the participants in the intervention group, which resulted in a large increase in motivation to receive the HPV vaccine among the participants. Similarly, a 3-day education program on decision-making, self-efficacy, and objective skills regarding HPV vaccination was given to the intervention group, which not only guided on the ways to receive financial and emotional support from family and friends, but also clarified the appointment process of HPV vaccination and other related issues of concern. The tailored factual information of this study covered almost all of the possible difficulties that the students might encounter after they decided to receive the HPV vaccine. Previous studies [43-45] showed that compared with awareness and attitudes toward health behaviors, the adoption of health behaviors usually had a certain time lag and needed to be gradually realized in long-term positive practice. In addition, when the scores of the intervention and control groups were compared, the intervention group scored better than the control group in each survey after the education program. The results of this study suggest that although the 7-day education program might not have been sufficient to significantly improve motivation and behavioral skills regarding HPV vaccination in the intervention group at the end of the study, it can maintain participants' motivation and behavioral skills at certain levels and prevent them from falling.

### Perceptions of the Barriers and Facilitators of Receiving the HPV Vaccine

About two-thirds of the participants in our study were willing to be vaccinated against HPV, which is higher than the proportion reported in previous studies [46-48], and 90.5% of them were likely to recommend the HPV vaccine to their friends or relatives, which suggests that most of them agreed with the benefits of the HPV vaccine, but had more concerns when it came to themselves, such as concerns about the safety, effectiveness, and possible side effects of the HPV vaccine [49]. In addition, about a quarter of the participants in the control group believed that a lack of knowledge about HPV-related issues was holding them back from receiving the HPV vaccine, which suggests that importance should be given to improving the knowledge and awareness of HPV vaccination among target populations. Additionally, the government should take efforts to make the HPV vaccine affordable and accessible for most college students; to ensure the supply of the domestically produced HPV vaccine, which is popular among female college students; and to promote the coverage of HPV vaccination by medical insurance or its inclusion in national immunization programs as soon as possible. Furthermore, the results of this study suggest that in order to promote HPV vaccination, media publicity of the HPV vaccine should be strengthened nationwide, HPV-related information should be added to the in-service education of doctors, and doctors should be encouraged to recommend the HPV vaccine to female individuals during the consultation process.

## Strengths and Limitations

This is the first randomized trial to study the effectiveness of an IMB model–guided education program to improve HPV vaccination among female college students in mainland China. The strengths of this study include its multicenter design, the inclusion of a representative and sufficient sample of the target population, and the application of a health behavioral theory. Owing to online education and participant management, our study was carried out as scheduled even during the period of the COVID-19 outbreak. This provided a unique opportunity to make clear statements about the feasibility and convenience of the study. In addition, most of the possible contaminations were avoided owing to home isolation of the participants during the study period. The quality control measures of the study can be considered strengths. They included a professional research design and pilot trial conducted by researchers with public health and epidemiological backgrounds, daily quizzes and regular reminders to improve the maintenance rate during follow-up, and logical questions to ensure the quality of the available data.

Several study limitations must be considered. First, the effectiveness of the study might have been compromised by cluster randomization and a mismatch in the number of participants in the intervention and control groups. In addition, although the IMB model–based intervention showed relatively desirable results, the adverse impacts of the COVID-19 pandemic could not be ignored, which included too much concern about getting COVID-19 and the possible trauma, as well as difficulties in making an appointment and getting vaccinated against HPV. Moreover, the participants were asked to self-report if they had received the HPV vaccine before and within the study; however, the accuracy of these self-reported data could not be verified through the community medical care system. Future studies with randomization at the individual level, a longer follow-up period, and regular repetition of education will allow better exploration of the effects of the intervention.

## Conclusions

Online education based on the IMB model can improve HPV-related knowledge, as well as the motivation, behavioral skills, and willingness regarding HPV vaccination among female college students in the short term; however, its long-term effects may need to be consolidated and maintained by repeated education programs in the future.

---

## Acknowledgments

This work was supported by the Asian Regional Special Cooperation Fund of the National Health Commission of the People’s Republic of China (BLXM01), and the Innovation Fund for Medical Sciences sponsored by the Chinese Academy of Medical Sciences (2019-I2M-2-005, 2017-I2M-B&R-17).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist.
[PDF File (Adobe PDF File), 2389 KB - jmir_v24i9e37848_app1.pdf ]

## References

1. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. The Lancet Global Health 2020 Feb;8(2):e191-e203. [doi: 10.1016/s2214-109x(19)30482-6]
2. Global Burden of Disease 2019 Cancer Collaboration, Kocarnik JM, Compton K, Dean FE, Fu W, Gaw BL, et al. . [doi: 10.1001/jamaoncol.2021.6987] [Medline: 34967848]
3. Protecting adolescents with HPV vaccine. World Health Organization. 2018. URL: https://www.who.int/news-room/feature-stories/detail/protecting-adolescents-with-hpv-vaccine [accessed 2022-09-10]
4. Zhao F, Tiggelaar SM, Hu S, Xu L, Hong Y, Niyazi M, et al. A multi-center survey of age of sexual debut and sexual behavior in Chinese women: suggestions for optimal age of human papillomavirus vaccination in China. Cancer Epidemiol 2012 Aug;36(4):384-390 [FREE Full text] [doi: 10.1016/j.canep.2012.01.009] [Medline: 22377277]
5. Wang R, Guo X, Wisman GBA, Schuuring E, Wang W, Zeng Z, et al. Nationwide prevalence of human papillomavirus infection and viral genotype distribution in 37 cities in China. BMC Infect Dis 2015 Jul 04;15(1):257 [FREE Full text] [doi: 10.1186/s12879-015-0998-5] [Medline: 26142044]
6. Deng C, Chen X, Liu Y. Human papillomavirus vaccination: coverage rate, knowledge, acceptance, and associated factors in college students in mainland China. Hum Vaccin Immunother 2021 Mar 04;17(3):828-835 [FREE Full text] [doi: 10.1080/21645515.2020.1797368] [Medline: 32873128]
7. Healthy China Action - Cancer prevention and treatment three-year action plan (2019-2022). National Health Commission of the People's Republic of China. 2019. URL: http://www.nhc.gov.cn/jkj/s5878/201909/2cb5dfb5d4f84f8881897e232b376b60.shtml [accessed 2022-09-10]
8. World Health Assembly adopts global strategy to accelerate cervical cancer elimination. World Health Organization. 2020. URL: https://www.who.int/news/item/19-08-2020-world-health-assembly-adopts-global-strategy-to-accelerate-cervical-cancer-elimination [accessed 2022-09-10]
9. China supports "Global Strategy to Accelerate Elimination of Cervical Cancer". The State Council the People's Republic of China. 2020. URL: http://www.gov.cn/xinwen/2020-12/10/content_5568770.htm [accessed 2022-09-10]
10. Pan X, Li R, Pan A, Larson H. Human papillomavirus vaccine approval in China: a major step forward but challenges ahead. The Lancet Infectious Diseases 2016 Dec;16(12):1322-1323. [doi: 10.1016/s1473-3099(16)30450-9]
11. Drolet M, Laprise J, Martin D, Jit M, Bénard É, Gingras G, et al. Optimal human papillomavirus vaccination strategies to prevent cervical cancer in low-income and middle-income countries in the context of limited resources: a mathematical modelling analysis. The Lancet Infectious Diseases 2021 Nov;21(11):1598-1610. [doi: 10.1016/s1473-3099(20)30860-4]

12. Shi W, Cheng X, Wang H, Zang X, Chen T. Cost-effectiveness of human papillomavirus vaccine in China: a systematic review of modelling studies. BMJ Open 2021 Dec 08;11(12):e052682 [FREE Full text] [doi: 10.1136/bmjopen-2021-052682] [Medline: 34880019]
13. Zou H, Wang W, Ma Y, Wang Y, Zhao F, Wang S, et al. How university students view human papillomavirus (HPV) vaccination: A cross-sectional study in Jinan, China. Hum Vaccin Immunother 2016 Aug 26;12(1):39-46 [FREE Full text] [doi: 10.1080/21645515.2015.1072667] [Medline: 26308701]
14. Wang S, Zhang S, Pan X, Ren Z, Yang C, Wang Z, et al. Human papillomavirus vaccine awareness, acceptability, and decision-making factors among Chinese college students. Asian Pac J Cancer Prev 2014 Apr 01;15(7):3239-3245 [FREE Full text] [doi: 10.7314/apjcp.2014.15.7.3239] [Medline: 24815477]
15. Zhang Y, Wang Y, Liu L, Fan Y, Liu Z, Wang Y, et al. Awareness and knowledge about human papillomavirus vaccination and its acceptance in China: a meta-analysis of 58 observational studies. BMC Public Health 2016 Mar 03;16(1):216 [FREE Full text] [doi: 10.1186/s12889-016-2873-8] [Medline: 26936076]
16. Prasetya H, Murti B, Anantanyu S, Syamsulhadi M. The effect of hypnosis on adherence to antituberculosis drugs using the health belief model. Int J Clin Exp Hypn 2018 Mar 30;66(2):211-227. [doi: 10.1080/00207144.2018.1421361] [Medline: 29601278]
17. Shamizadeh T, Jahangiry L, Sarbakhsh P, Ponnet K. Social cognitive theory-based intervention to promote physical activity among prediabetic rural people: a cluster randomized controlled trial. Trials 2019 Feb 04;20(1):98 [FREE Full text] [doi: 10.1186/s13063-019-3220-z] [Medline: 30717779]
18. Hosseini SS, Shamsi M, Khorsandi M, Moradzadeh R. The effect of educational program based on theory of planned behavior on promoting retinopathy preventive behaviors in patients with type 2 diabetes: RCT. BMC Endocr Disord 2021 Jan 14;21(1):17 [FREE Full text] [doi: 10.1186/s12902-021-00680-2] [Medline: 33446157]
19. Perez GK, Cruess DG, Strauss NM. A brief information-motivation-behavioral skills intervention to promote human papillomavirus vaccination among college-aged women. Psychology research and behavior management 2016 Oct;Volume 9:285-296. [doi: 10.2147/prbm.s112504]
20. Xiao X, Lee DKL, Wong RM, Borah P. The impact of theory in HPV vaccination promotion research: A systematic review and meta-analysis. Am J Health Promot 2021 Sep 05;35(7):1002-1014. [doi: 10.1177/08901171211012524] [Medline: 33949203]
21. Zhang X, Liu CR, Wang ZZ, Ren ZF, Feng XX, Ma W, et al. Effect of a school-based educational intervention on HPV and HPV vaccine knowledge and willingness to be vaccinated among Chinese adolescents : a multi-center intervention follow-up study. Vaccine 2020 Apr 29;38(20):3665-3670 [FREE Full text] [doi: 10.1016/j.vaccine.2020.03.032] [Medline: 32245644]
22. Chang IJ, Huang R, He W, Zhang S, Wang S, Zhao F, et al. Effect of an educational intervention on HPV knowledge and vaccine attitudes among urban employed women and female undergraduate students in China: a cross-sectional study. BMC Public Health 2013 Oct 02;13(1):916 [FREE Full text] [doi: 10.1186/1471-2458-13-916] [Medline: 24088392]
23. Kwan TT, Tam K, Lee PW, Chan KK, Ngan HY. The effect of school-based cervical cancer education on perceptions towards human papillomavirus vaccination among Hong Kong Chinese adolescent girls. Patient Educ Couns 2011 Jul;84(1):118-122. [doi: 10.1016/j.pec.2010.06.018] [Medline: 20650592]
24. Inner Mongolia banner to offer free HPV shots. China Daily. 2020. URL: https://enapp.chinadaily.com.cn/a/202008/11/AP5f31eea1a3109daf6c775978.html [accessed 2022-09-10]
25. Si M, Su X, Jiang Y, Qiao Y, Liu Y. Interventions to improve human papillomavirus vaccination among Chinese female college students: study protocol for a randomized controlled trial. BMC Public Health 2019 Nov 21;19(1):1546 [FREE Full text] [doi: 10.1186/s12889-019-7903-x] [Medline: 31752785]
26. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
27. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
28. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992 May;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
29. Fisher WA. Understanding human papillomavirus vaccine uptake. Vaccine 2012 Nov 20;30 Suppl 5:F149-F156. [doi: 10.1016/j.vaccine.2012.04.107] [Medline: 23199958]
30. Chen P, Chun-rong L, Xi Z, Jing L. Knowledge and attitudes toward HPV and its vaccines among parents of middle school students in Chengdu. Modern Preventive Medicine 2018;45:299-302 [FREE Full text]
31. Kim HW. Knowledge about human papillomavirus (HPV), and health beliefs and intention to recommend HPV vaccination for girls and boys among Korean health teachers. Vaccine 2012 Aug 03;30(36):5327-5334. [doi: 10.1016/j.vaccine.2012.06.040] [Medline: 22749602]
32. Kahn JA, Rosenthal SL, Hamann T, Bernstein DI. Attitudes about human papillomavirus vaccine in young women. Int J STD AIDS 2003 May 25;14(5):300-306. [doi: 10.1258/095646203321605486] [Medline: 12803935]

33. Forster A, McBride K, Davies C, Stoney T, Marshall H, McGeechan K, et al. Development and validation of measures to evaluate adolescents' knowledge about human papillomavirus (HPV), involvement in HPV vaccine decision-making, self-efficacy to receive the vaccine and fear and anxiety. Public Health 2017 Jun;147:77-83 [FREE Full text] [doi: 10.1016/j.puhe.2017.02.006] [Medline: 28404501]
34. DiClemente RJ, Murray CC, Graham T, Still J. Overcoming barriers to HPV vaccination: A randomized clinical trial of a culturally-tailored, media intervention among African American girls. Hum Vaccin Immunother 2015 Sep 17;11(12):2883-2894 [FREE Full text] [doi: 10.1080/21645515.2015.1070996] [Medline: 26378650]
35. Gerend MA, Shepherd JE. Predicting human papillomavirus vaccine uptake in young adult women: comparing the health belief model and theory of planned behavior. Ann Behav Med 2012 Oct 1;44(2):171-180 [FREE Full text] [doi: 10.1007/s12160-012-9366-5] [Medline: 22547155]
36. Guvenc G, Seven M, Akyuz A. Health belief model scale for human papilloma virus and its vaccination: Adaptation and psychometric testing. J Pediatr Adolesc Gynecol 2016 Jun;29(3):252-258. [doi: 10.1016/j.jpag.2015.09.007] [Medline: 26409648]
37. Si M, Jiang Y, Su X, Wang W, Zhang X, Gu X, et al. Willingness to accept human papillomavirus vaccination and its influencing factors using information-motivation-behavior skills model: A cross-sectional study of female college freshmen in Mainland China. Cancer Control 2021 Oct 11;28:10732748211032899 [FREE Full text] [doi: 10.1177/10732748211032899] [Medline: 34634207]
38. Chi X, Hawk ST, Winter S, Meeus W. The effect of comprehensive sexual education program on sexual health knowledge and sexual attitude among college students in Southwest China. Asia Pac J Public Health 2015 Mar 14;27(2):NP2049-NP2066. [doi: 10.1177/1010539513475655] [Medline: 23417908]
39. Starosta AJ, Cranston E, Earleywine M. Safer sex in a digital world: A Web-based motivational enhancement intervention to increase condom use among college women. J Am Coll Health 2016;64(3):184-193. [doi: 10.1080/07448481.2015.1107835] [Medline: 26700487]
40. Icenogle G, Cauffman E. Adolescent decision making: A decade in review. J Res Adolesc 2021 Dec 24;31(4):1006-1022. [doi: 10.1111/jora.12608] [Medline: 34820945]
41. Wang C, Cheng Z, Yue X, McAleer M. Risk management of COVID-19 by universities in China. JRFM 2020 Feb 19;13(2):36. [doi: 10.3390/jrfm13020036]
42. Ogundele OA, Omotoso AA, Fagbemi AT. COVID-19 outbreak: a potential threat to routine vaccination programme activities in Nigeria. Hum Vaccin Immunother 2021 Mar 04;17(3):661-663 [FREE Full text] [doi: 10.1080/21645515.2020.1815490] [Medline: 32991237]
43. Bennett AT, Patel DA, Carlos RC, Zochowski MK, Pennewell SM, Chi AM, et al. Human papillomavirus vaccine uptake after a tailored, online educational intervention for female university students: A randomized controlled trial. J Womens Health (Larchmt) 2015 Nov;24(11):950-957 [FREE Full text] [doi: 10.1089/jwh.2015.5251] [Medline: 26488269]
44. Henrikson NB, Zhu W, Baba L, Nguyen M, Berthoud H, Gundersen G, et al. Outreach and reminders to improve human papillomavirus vaccination in an integrated primary care system. Clin Pediatr (Phila) 2018 Nov 13;57(13):1523-1531. [doi: 10.1177/0009922818787868] [Medline: 30003794]
45. Liu C, Liang H, Zhang X, Pu C, Li Q, Li Q, et al. Effect of an educational intervention on HPV knowledge and attitudes towards HPV and its vaccines among junior middle school students in Chengdu, China. BMC Public Health 2019 May 02;19(1):488 [FREE Full text] [doi: 10.1186/s12889-019-6823-0] [Medline: 31046722]
46. You D, Han L, Li L, Hu J, Zimet GD, Alias H, et al. Human papillomavirus (HPV) vaccine uptake and the willingness to receive the HPV vaccination among female college students in China: A multicenter study. Vaccines (Basel) 2020 Jan 16;8(1):31 [FREE Full text] [doi: 10.3390/vaccines8010031] [Medline: 31963370]
47. Ma Y, Wang C, Liu F, Lian G, Li S, He Q, et al. Human papillomavirus vaccination coverage and knowledge, perceptions and influencing factors among university students in Guangzhou, China. Hum Vaccin Immunother 2021 Oct 03;17(10):3603-3612 [FREE Full text] [doi: 10.1080/21645515.2021.1927411] [Medline: 34156312]
48. Lin Y, Lin Z, He F, Hu Z, Zimet GD, Alias H, et al. Factors influencing intention to obtain the HPV vaccine and acceptability of 2-, 4- and 9-valent HPV vaccines: A study of undergraduate female health sciences students in Fujian, China. Vaccine 2019 Oct 16;37(44):6714-6723. [doi: 10.1016/j.vaccine.2019.09.026] [Medline: 31548016]
49. Pomares TD, Buttenheim AM, Amin AB, Joyce CM, Porter RM, Bednarczyk RA, et al. Association of cognitive biases with human papillomavirus vaccine hesitancy: a cross-sectional study. Hum Vaccin Immunother 2020 May 03;16(5):1018-1023 [FREE Full text] [doi: 10.1080/21645515.2019.1698243] [Medline: 31859593]

## Abbreviations

**GEE:** generalized estimating equation
**HPV:** human papillomavirus
**IMB:** information-motivation-behavioral skills

*Edited by R Kukafka; submitted 13.03.22; peer-reviewed by R Mpofu, ER Khalilian; comments to author 03.06.22; revised version received 28.07.22; accepted 06.09.22; published 30.09.22.*

*Please cite as:*
*Si M, Su X, Jiang Y, Wang W, Zhang X, Gu X, Ma L, Li J, Zhang S, Ren Z, Liu Y, Qiao Y*
*An Internet-Based Education Program for Human Papillomavirus Vaccination Among Female College Students in Mainland China: Application of the Information-Motivation-Behavioral Skills Model in a Cluster Randomized Trial*
*J Med Internet Res 2022;24(9):e37848*
*URL: https://www.jmir.org/2022/9/e37848*
*doi:10.2196/37848*
*PMID:36178723*

Original Paper

# The Impact of a Clinical Decision Support System for Addressing Physical Activity and Healthy Eating During Smoking Cessation Treatment: Hybrid Type I Randomized Controlled Trial

Nadia Minian[1,2,3,4,5], PhD; Mathangee Lingam[1], BSc, MPH; Rahim Moineddin[2], PhD; Kevin E Thorpe[6,7], MMath; Scott Veldhuizen[1], PhD; Rosa Dragonetti[1,2], MSc; Laurie Zawertailo[1,5], PhD; Valerie H Taylor[8], MD, PhD; Margaret Hahn[4,9,10,11], MD, PhD; Wayne K deRuiter[1], PhD; Osnat C Melamed[1,2], MSc, MD; Peter Selby[1,2,6,9], MBBS, MHSc

[1]Nicotine Dependence Service, Centre for Addiction and Mental Health, Toronto, ON, Canada
[2]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[3]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada
[4]Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada
[5]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, Canada
[6]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[7]Applied Health Research Centre, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada
[8]Department of Psychiatry, University of Calgary, Calgary, ON, Canada
[9]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[10]Schizophrenia Division, Centre for Addiction and Mental Health, Toronto, ON, Canada
[11]Banting and Best Diabetes Centre, University of Toronto, Toronto, ON, Canada

**Corresponding Author:**
Peter Selby, MBBS, MHSc
Nicotine Dependence Service
Centre for Addiction and Mental Health
1025 Queen Street West
Toronto, ON, M6J 1H1
Canada
Phone: 1 416 535 8501 ext 36859
Email: peter.selby@camh.ca

## Abstract

**Background:** People who smoke have other risk factors for chronic diseases, such as low levels of physical activity and poor diet. Clinical decision support systems (CDSSs) might help health care practitioners integrate interventions for diet and physical activity into their smoking cessation programming but could worsen quit rates.

**Objective:** The aims of this study are to assess the effects of the addition of a CDSS for physical activity and diet on smoking cessation outcomes and to assess the implementation of the study.

**Methods:** We conducted a pragmatic hybrid type I effectiveness-implementation trial with 232 team-based primary care practices in Ontario, Canada, from November 2019 to May 2021. We used a 2-arm randomized controlled trial comparing a CDSS addressing physical activity and diet to treatment as usual and used the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework to measure implementation outcomes. The primary outcome was self-reported 7-day tobacco abstinence at 6 months.

**Results:** We enrolled 5331 participants in the study. Of these, 2732 (51.2%) were randomized to the intervention group and 2599 (48.8%) to the control group. At the 6-month follow-up, 29.7% (634/2137) of respondents in the intervention arm and 27.3% (552/2020) in the control arm reported abstinence from tobacco. After multiple imputation, the absolute group difference was 2.1% (95% CI −0.5 to 4.6; $F_{1,1000.42}$=2.43; *P*=.12). Mean exercise minutes changed from 32 (SD 44.7) to 110 (SD 196.1) in the intervention arm and from 32 (SD 45.1) to 113 (SD 195.1) in the control arm (group effect: B=−3.7 minutes; 95% CI −17.8 to 10.4; *P*=.61). Servings of fruit and vegetables changed from 2.64 servings to 2.42 servings in the intervention group and from 2.52 servings to 2.45 servings in the control group (incidence rate ratio for intervention group=0.98; 95% CI 0.93-1.02; *P*=.35).

**Conclusions:** A CDSS for physical activity and diet may be added to a smoking cessation program without affecting the outcomes. Further research is needed to improve the impact of integrated health promotion interventions in primary care smoking cessation programs.

**Trial Registration:** ClinicalTrials.gov NCT04223336 https://www.clinicaltrials.gov/ct2/show/NCT04223336

**International Registered Report Identifier (IRRID):** RR2-10.2196/19157

***(J Med Internet Res 2022;24(9):e37900)*** doi:10.2196/37900

**KEYWORDS**

smoking cessation; physical activity; healthy eating; clinical decision support system; Canada; diet; intervention; smoking; primary care; program; treatment; clinical decision support; health behavior

# *Introduction*

## Background

Smoking, low levels of physical activity, and poor dietary habits are highly prevalent and the three leading behavioral causes of death worldwide [1-4]. The concurrence of these risk factors [5-8] compounds the risks of developing chronic diseases [5,9]. Behavioral interventions by health care practitioners addressing these risk factors are potentially cost-effective [10,11]. Furthermore, improvements in one behavior can positively impact other risky behaviors [5,12,13]. For example, increasing physical activity can help reduce acute cravings and withdrawal symptoms when quitting smoking [14-17]. While the link between improving dietary habits and smoking cessation is less clear, improving dietary habits may prevent some postcessation weight gain. This can be a barrier to quitting smoking and maintaining abstinence [18-20]. Given these relationships, it is important to adopt a holistic approach to addressing risk behaviors [21].

Clinical decision support systems (CDSSs) are a promising resource to effectively support health care practitioners with the delivery of integrated evidence-based interventions to their patients [22-24]. A CDSS is an electronic application that can synthesize complex patient-specific information and present tailored recommendations to health care practitioners in real time [22-24]. They are frequently used in health care settings to help improve adherence to clinical guidelines, reduce treatment errors, and improve preventive care [25-30]. Moreover, a CDSS collects relevant data from different sources and presents them to the user in a central and easily accessible format. This is associated with improved efficiency and alleviates time burden during treatment planning [31]. Since many primary care offices use electronic medical records [32], CDSSs are also well-suited for seamless integration into existing workflows and allows for rapid and widespread scalability. We demonstrated that the addition of a CDSS in a smoking cessation program increased the likelihood that patients with at-risk drinking accepted an educational resource to reduce or abstain from alcohol consumption [33].

## Objective

The Smoking Treatment for Ontario Patients (STOP) program is a province-wide initiative that works in partnership with primary care settings across Ontario to provide tobacco users with up to 26 weeks of behavioral counseling and no-cost nicotine replacement therapy for smoking cessation. Health care practitioners at these organizations use the STOP portal, a web-based data collection and treatment management tool, to enroll their patients into the STOP program. However, an analysis of former STOP participants showed that 62% and 96% of STOP participants reported being below the Canadian guidelines for physical activity [34] and fruit and vegetable consumption [35], respectively. The STOP portal currently has a built-in CDSS to guide health care practitioners with addressing depressive symptoms and at-risk alcohol use as part of smoking cessation treatment. Scaling up the STOP portal to incorporate an additional CDSS that encourages practitioners to address patients' physical activity and diet as part of the overall smoking cessation treatment could provide an opportunity to improve smoking cessation rates. However, there are several potential risks and limitations with implementing CDSS, including alert fatigue and disruptions to the current workflow of health care practitioners [24]. While the CDSS is designed to streamline processes and promote integration, it may still have unintended negative consequences. For example, the CDSS may require clinicians to engage in additional steps to input data, which can disrupt their clinical workflow and reduce clinician time to treat the patient's presenting complaint. The CDSS may also not be relevant for all patient populations or clinical encounters, and this could inadvertently introduce bias in treatment [36]. Since tobacco use is correlated with the largest reductions in health-adjusted life expectancy [1], a key part of the implementation of the CDSS for other modifiable risk behaviors should be to ensure that it does not negatively affect the likelihood of patients quitting smoking.

The aims of this study were to (1) assess whether adding a CDSS for physical activity and diet to a smoking cessation program positively or negatively affects smoking cessation outcomes and (2) assess the implementation of the study using the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework [37]. The detailed protocol is described elsewhere [38]. In this manuscript, we report the quantitative findings.

# *Methods*

## Study Design

We conducted a pragmatic, hybrid type I effectiveness and implementation trial [39], which allowed for the simultaneous testing of intervention effectiveness and implementation feasibility in real-world settings. Health care practitioners in primary care settings often report barriers to offering comprehensive preventive health services, including not having enough time, skills, knowledge, or resources [40]. As a result, it is important to examine the effectiveness of the CDSS on patients' behavior outcomes, as well as how the CDSS impacts health care practitioners' ability to provide treatment to their patients. Examining both effectiveness and implementation outcomes can help to provide valuable insights into the uptake of the CDSS and provide context for any observed results. We measured the effectiveness of the CDSS in a smoking cessation program using a 2-arm randomized controlled trial comparing a physical activity and diet CDSS directed at practitioners (intervention) with treatment as usual (control). The RE-AIM framework [37] was used to measure implementation outcomes in the intervention group.

## Setting and Location

The trial was operationalized in team-based primary care practices in Ontario, Canada (family health teams [n=153], community health centers [n=61], and nurse practitioner–led clinics [n=18]), implementing the STOP program as of November 29, 2019.

## Preimplementation Measures

To better equip health care practitioners to implement the intervention, we undertook several knowledge translation initiatives based on the principles of the Interactive Systems Framework (ISF) for dissemination and implementation [41]. The ISF comprises three interacting systems that facilitate the implementation of research in real-world practice: delivery systems, synthesis and translation systems, and support systems [41]. In this study, the health care practitioners at the primary clinics acted as the delivery system. The ISF's synthesis and translation systems were addressed by engaging STOP program participants in the cocreation of health behavior change messages and a self-monitoring resource for tracking their health behaviors [42]. As part of the ISF support system, we provided health care practitioners with training (via an interactive webinar [43]) around evidence-based recommendations for addressing physical activity and fruit and vegetable consumption as part of a smoking cessation treatment program [44].

## Extenuating Circumstances

Most studies were conducted during the COVID-19 pandemic. The COVID-19 pandemic state of emergency (SOE) was announced in Ontario on March 17, 2020 [45], which was approximately 4 months into the study. During this period, many of the primary care settings that implemented the STOP program transitioned to offering phone- and video-based appointments (eg, via phone or video) to their patients [45]. Consequently, practitioners had to virtually communicate any recommendations generated by the CDSS for physical activity and diet to patients. Therefore, we also examined how the pandemic may have affected the delivery of the intervention.

## Participants

Eligible participants were treatment-seeking cigarette smokers who enrolled in the STOP program at one of the partnering primary care settings and reported baseline physical activity levels [34] and fruit and vegetable consumption levels [35] that were lower than the national guidelines. Low levels of physical activity were defined as engaging in less than 150 minutes of moderate to vigorous exercise per week (Canadian Physical Activity Guidelines) [34]. Low levels of fruit and vegetable consumption were defined as consuming less than 7 servings (female) or 8 servings (male) of fruit and vegetables daily (2007 Canada's Food Guide) [35]. Participants were also required to be English-speaking and provide at least one piece of contact information (email address or phone number) so that the study team could conduct follow-up surveys at 6 months following enrollment into the STOP program. Participants were enrolled in the STOP program through self-referral or practitioner referral. Assessment of whether a participant met the eligibility criteria (as listed above) for the study was determined using the patient's self-reported responses to the corresponding questions in the STOP program's enrollment survey. The enrollment survey was completed using the STOP program's web-based portal (STOP portal). The eligibility criteria remained the same throughout the study and were not affected by changes in response to the COVID-19 pandemic.

## Treatment Arms

### *Intervention Arm*

The intervention was a CDSS that alerted health care practitioners if their patient reported low levels of physical activity and fruit and vegetable consumption and provided recommendations for behavior interventions. The recommendations in the CDSS were based on the literature on the most effective types of behavior change techniques [44]. The CDSS first prompted the practitioners to provide a brief risk communication intervention for physical activity and fruit and vegetable consumption. The risk communication involved sharing information about the risk behavior and discussing how it would affect both the patient's health and smoking cessation treatment. The practitioners were then prompted to provide (print or via email) the patient with a self-monitoring resource for these risk behaviors. The self-monitoring resource was a 1-page paper-based weekly tracking sheet that patients could use to record their smoking, physical activity, and fruit and vegetable consumption.

### *Control Arm*

In the control arm, the CDSS did not alert practitioners to whether their patients reported low levels of physical activity and fruit and vegetable consumption and did not provide practitioner recommendations to address these risk behaviors. Health care practitioners experienced the STOP portal as usual, which includes a CDSS for depressive symptoms and alcohol use. Although the CDSS was not available in the control group, practitioners were not prevented from addressing physical activity and diet with their patients if they deemed it clinically appropriate. We did not track whether the practitioner provided any counseling to the control group.

## Outcomes

We used the RE-AIM framework to structure and interpret the study outcomes. The components of the RE-AIM framework are reach, effectiveness, adoption, implementation, and maintenance.

### *Reach*

The reach of intervention was assessed by examining the changes in the proportion of enrollments that were recorded as having been completed directly on the portal, before and after the CDSS for physical activity and fruit and vegetable consumption were introduced.

Health care practitioners can administer the STOP baseline questionnaire using the portal or on paper. Given that the CDSS is only available to practitioners when they conduct the questionnaire using the portal, any decrease in this proportion could be an indication that practitioners were avoiding the intervention by switching to a paper enrollment.

### *Effectiveness*

The primary outcome of interest for this study was self-reported smoking cessation (7-day point prevalence abstinence) at the 6-month follow-up following enrollment. This outcome was measured by a response of “No” to the following question: “Have you smoked a cigarette, even a puff, in the last 7 days?” Research comparing self-reported smoking status with a biochemical assessment of smoking has found that they are highly correlated [46-49]. The secondary outcomes were self-reported changes in physical activity levels and fruit and vegetable consumption levels between baseline to 6-month follow-up. The Exercise Vital Signs Screener [50] was adapted to assess changes in physical activity. This screener has been validated [50] and consists of two questions: “On average, how many days per week do you engage in moderate-to-strenuous (vigorous) exercise (like a brisk walk)?” “On these days, for how many minutes do you typically exercise at this level?” The responses to these 2 questions were multiplied to produce the total minutes per week of moderate-to-vigorous exercise. Fruit and vegetable consumption was measured using a single question: “In a typical day, how many total servings of fruits and vegetables do you eat? (One serving is 1/2 cup of fresh, frozen, or canned fruits or vegetables, or 1/2 cup of 100% juice. Please DO NOT include potatoes).”

At baseline (time of enrollment), these self-report questions were administered by a STOP practitioner. At the 6-month follow-up, we collected responses to these questions via phone, email, or during a visit to a practitioner. The threshold for significance for the primary and secondary outcomes was $P<.05$. This is the established standard in empirical research for determining statistical significance despite its limitations [51].

### *Adoption*

Given that the intervention has several components and was personalized to meet the needs of the participants, we assessed how often each component was adopted. Specifically, we examined the proportion of participants in the intervention group who were offered a self-monitoring resource for physical activity and fruit and vegetable consumption by their health care practitioner when appropriate.

### *Implementation*

To acquire an understanding of the degree of fidelity at which the intervention was implemented, we examined how many participants accepted the self-monitoring resources. The intervention was classified as fully implemented when eligible participants received the corresponding self-monitoring resources. The patient only had the opportunity to accept the resource if the practitioner made the decision to offer it to the patient.

### *Maintenance*

To determine whether efforts to provide the intervention changed over time, we calculated the proportion of eligible enrollees in each month who were offered the self-monitoring resources. To examine the sustainability of the smoking cessation intervention, we calculated the proportion of participants who had stopped smoking at both the 6-month and 12-month follow-ups.

In our protocol, we had only outlined the patient-level outcome for our maintenance outcome. Upon reflection, we also included the outcomes at the setting level.

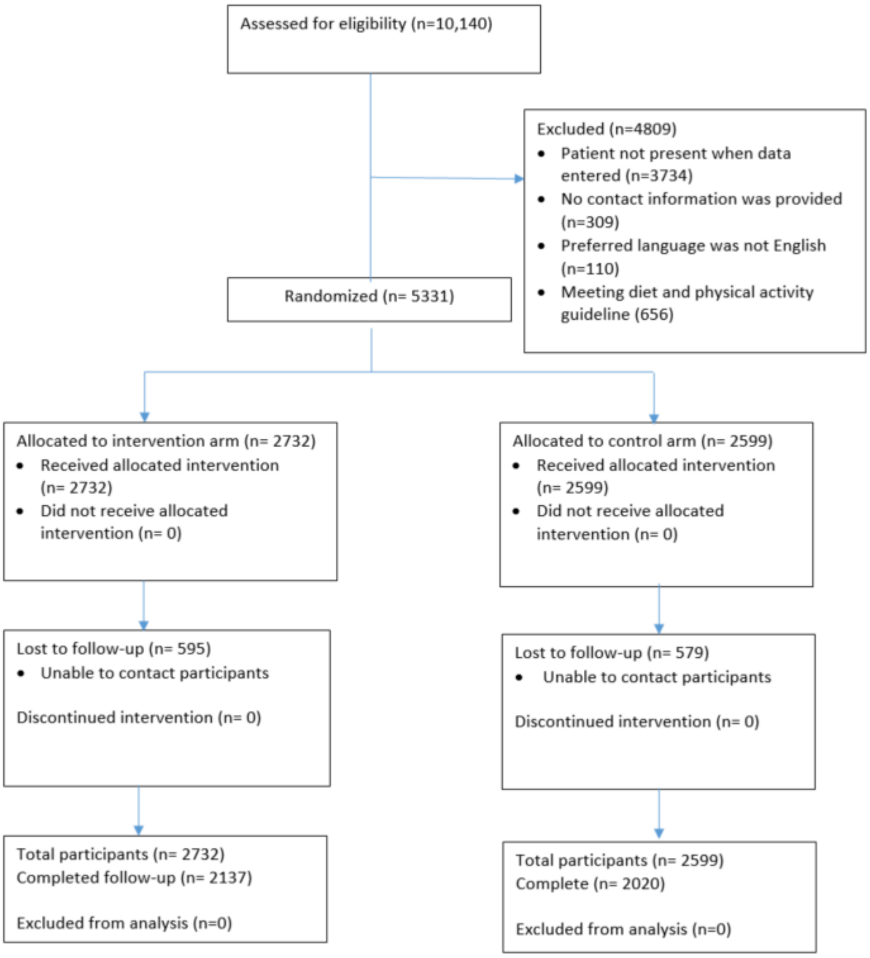
## Sample Size

We determined the sample size required to detect clinically meaningful differences in our primary outcome was 3998 participants (1999 per group). For this sample size calculation, the effect size was an absolute difference in proportions of 0.04, the standard in smoking cessation for clinical significance [52]. On the basis of the past STOP program 6-month follow-up data, the proportion of individuals who will quit at 6 months was estimated to be 0.26. The power was set to 80%, and $\alpha$ was set to .05.

On the basis of the STOP program’s follow-up completion rate in the years before this trial, we anticipated a loss to follow-up of 25%. This increased the necessary total sample size to 5331 (2666 per group).

As the randomization and intervention pathways were built into the STOP portal as part of the enrollment survey, we analyzed data when all participants in both arms had completed their 6-month follow-up. The analysis sample (Figure 1) included 4157 (control: n=2020, 48.6%; intervention: n=2137, 51.4%) participants who responded at the 6-month follow-up, as well as an additional 1174 participants who provided baseline and clinical data only (total sample, n=5331; control: n=2599, 48.8%; intervention: n=2732, 51.2%). Follow-up rates were 78% in both arms.

**Figure 1.** CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the number participants allocated to intervention and control group and included in our primary and secondary data analyses.



## Randomization

At the time of enrollment, we used a simple 1-to-1 randomization to assign eligible STOP participants to the intervention or control arm. The random allocation sequence was generated and implemented automatically by the STOP portal. The STOP portal performed the randomization using a built-in random number generator that determined, at the time of enrollment, which group the patient will be randomized to. The portal then automatically assigned the patient to the corresponding group. The practitioner was unable to override the allocation to the treatment arms. As health care practitioners treated both intervention and control patients, they were not blinded to the patient's treatment allocation. However, patients were blinded to the study arm assignment.

## Statistical Analysis

Our analysis followed the principle of intention-to-treat, in which all participants were analyzed in the groups to which they were randomized. As specified by the study protocol, we tested differences in our primary outcome (the proportion of participants abstinent from cigarettes at the 6-month follow-up) using a chi-square test and differences in our secondary outcomes by regressing the follow-up measure on the baseline value and the study group. We used linear regression for the total weekly minutes of physical activity and negative binomial regression for the daily servings of fruits and vegetables. For these secondary outcomes, we only included participants who received each specific pathway (diet: control, n=2529 [47%]; intervention, n=2618 [49%]; exercise, control; n=1747 [33%]; intervention, n=1879 [35%]).

To minimize possible bias, we retained participants with missing baseline variables or follow-up data in our analysis. To address missing data, we used multivariate imputation by chained equations. In our multiple imputation models, we included variables capturing demographic and socioeconomic characteristics, heaviness of smoking, health conditions, self-rated importance of quitting and confidence in the ability to quit, the type of nicotine replacement therapy provided at the initial clinical contact, the number of clinical visits attended, and quit status at other time points (the most recent clinical visit and the 3-month and 12-month follow-ups). A total of 50 imputed data sets were generated. We used the mitools package in R 3.6 [53,54] to produce a chi-squared statistic with multiple imputation data sets. This produced an F-distributed D2 statistic.

### *Per-Protocol Analysis*

The intervention comprised an automated alert to practitioners; not all patients in the intervention arm were offered the self-monitoring resources and not all patients who were offered the self-monitoring resources accepted them. To explore differences associated with the offering and acceptance of the resources, we conducted a secondary analysis in which we divided participants into four groups: the control arm, participant with whom practitioners conducted a risk communication discussion and offered the self-monitoring resources for these behaviors, participants receiving neither a risk communication discussion nor an offer of resources, and participants receiving either a discussion or an offer but not both. These analyses were similar to those described earlier but with 4 groups. If the participants were targeted with both interventions, we used the higher intensity level to determine their group.

### *Analysis of Reach Outcome*

The CDSS operated only when practitioners reported that they were completing the baseline survey in the presence of the patient. To test whether the implementation of CDSS changed this reporting (as might be expected if clinicians were avoiding the pathway), we fit a piecewise logistic regression model of change over time, with a random intercept for the clinic. The COVID-19 pandemic also began during data collection; as a result, we fit a 3-part spline with two indicator variables, one for the beginning of the study and one for the beginning of the pandemic.

### *Sensitivity Analyses*

To assess the effects of our treatment of missing data, we conducted a complete-case sensitivity analysis for our primary outcome by computing a chi-square statistic for the association between group and smoking cessation. To test for an effect of the COVID-19 pandemic on the effects of the study intervention, we fit an additional logistic regression model such as an interaction between the study arm and an indicator variable that was 1 if the enrollment took place after the SOE declaration and 0 otherwise.

## Ethics Approval

This study was approved by the Research Ethics Board of the Centre for Addiction and Mental Health (REB#:119-2018) and is registered at ClinicalTrials.gov (NCT04223336).

# *Results*

## Overview

Between November 2019 and October 2020, we enrolled 5331 patients with low levels of fruit and vegetable consumption or physical activity. Figure 1 provides a breakdown of the number of participants allocated to the intervention and control groups and included in our primary and secondary data analyses. We randomized 2599 (48.8%) participants to the control group and 2732 (51.2%) to the intervention group.

There were no major differences in the baseline characteristics of participants in the intervention and control groups (Table 1). Nearly all participants were daily smokers and over half have attempted to quit at least once in the year before their STOP enrollment. A little over a third of the participants reported being currently employed, and a quarter of the participants reported a household income below CAD $40,000 (USD $30,379.80).

In the intervention group, 853 (31%) participants reported low levels of fruit and vegetable consumption only, 114 (4%) reported low physical activity levels only, and 1765 (65%) reported both low levels of fruit and vegetable consumption and low physical activity levels. In the control group, 852 (33%) participants reported low fruit or vegetable consumption only, 70 (3%) reported low physical activity levels only, and 1677 (65%) reported both.

**Table 1.** Baseline characteristics of participants in study sample (n=5331).

| Baseline characteristics | Intervention (n=2732) | Control (n=2599) | Total missing, n (%) | *P* value |
|---|---|---|---|---|
| Age (years), mean (SD) | 52.9 (13.9) | 53.2 (13.7) | 1 (<1) | .34 |
| Sex (female), n (%) | 1464 (54) | 1321 (51) | 1 (<1) | .09 |
| First Nations, Metis, Inuit, n (%) | 174 (7) | 170 (7) | 154 (3) | .78 |
| High school diploma, n (%) | 1925 (75) | 1800 (74) | 321 (6) | .33 |
| Employed in past week, n (%) | 982 (36) | 882 (35) | 82 (2) | .34 |
| Household income CAD ≥$40,000 (US $30,379.80), n (%) | 377 (25) | 371 (25) | 2357 (44) | .78 |
| **Smoking** | | | | |
| Daily smoker, n (%) | 2576 (94) | 2430 (94) | 3 (<1) | .24 |
| Willing to set quit date, n (%) | 1905 (70) | 1808 (70) | 0 (<1) | .92 |
| Importance of quitting, mean (SD) | 9.2 (1.2) | 9.2 (1.2) | 22 (<1) | .46 |
| Confidence in ability to quit, mean (SD) | 7.5 (2.0) | 7.4 (2.0) | 40 (<1) | .26 |
| **Past-year quit attempts** | | | | .77 |
| None, n (%) | 1273 (47) | 1235 (48) | N/A[a] | |
| ≥1, n (%) | 1441 (53) | 1346 (52) | N/A | |
| Missing | N/A | N/A | 36 (1) | |
| **BMI (kg/m$^2$)** | | | | .79 |
| **Underweight (<18.5), n (%)** | 77 (3) | 70 (3) | N/A | |
| Normal (≥18.5 and <25), n (%) | 804 (32) | 745 (31) | N/A | |
| Overweight (≥25 and <30), n (%) | 825 (33) | 785 (33) | N/A | |
| Obese (≥30), n (%) | 786 (32) | 780 (33) | N/A | |
| Missing | N/A | N/A | 459 (9) | |
| At-risk drinking (AUDIT-C[b]), n (%) | 967 (36) | 899 (35) | 82 (2) | .50 |
| Low fruit or vegetable consumption levels, n (%) | 2618 (98) | 2529 (99) | 106 (2) | .01 |
| Low physical activity levels, n (%) | 1879 (70) | 1747 (69) | 109 (2) | .34 |
| At risk of depressive symptoms (PHQ-2[c] score ≥3), n (%) | 396 (15) | 365 (15) | 184 (3) | .62 |
| **Lifetime history of physical comorbid conditions** | | | | |
| Hypertension, n (%) | 924 (34) | 922 (36) | 68 (1) | .18 |
| High cholesterol, n (%) | 889 (33) | 875 (35) | 125 (2) | .34 |
| Heart disease, n (%) | 393 (15) | 376 (15) | 88 (2) | .90 |
| Stroke, n (%) | 149 (6) | 140 (5) | 61 (1) | .94 |
| Diabetes, n (%) | 437 (16) | 435 (17) | 73 (1) | .47 |
| COPD[d], n (%) | 808 (30) | 734 (29) | 152 (3) | .32 |
| Rheumatoid arthritis, n (%) | 199 (7) | 200 (8) | 140 (3) | .56 |
| Chronic pain, n (%) | 1050 (39) | 1002 (39) | 66 (1) | .91 |
| Cancer, n (%) | 264 (10) | 245 (10) | 73 (1) | .76 |
| **Lifetime history of psychiatric comorbid conditions** | | | | |
| Depression, n (%) | 1149 (43) | 1056 (41) | 78 (1) | .28 |
| Anxiety, n (%) | 1191 (44) | 1096 (43) | 82 (2) | .31 |
| Schizophrenia, n (%) | 78 (3) | 86 (3) | 164 (3) | .34 |
| Bipolar disorder, n (%) | 184 (7) | 166 (7) | 94 (2) | .67 |

| Baseline characteristics | Intervention (n=2732) | Control (n=2599) | Total missing, n (%) | *P* value |
|---|---|---|---|---|
| **Lifetime history of substance use disorder** | | | | |
| Drug use disorder, n (%) | 252 (9) | 214 (8) | 79 (1) | .21 |
| Alcohol use disorder, n (%) | 284 (11) | 253 (10) | 74 (1) | .39 |
| **Organization type** | | | | .05 |
| Family health team, n (%) | 2026 (74) | 1856 (71) | N/A | |
| Community health center, n (%) | 622 (23) | 666 (26) | N/A | |
| Nurse practitioner–led clinic, n (%) | 84 (3) | 77 (3) | N/A | |

[a] N/A: not applicable.

[b] AUDIT-C: Alcohol Use Disorders Identification Test.

[c] PHQ-2: Patient Health Questionnaire.

[d] COPD: chronic obstructive pulmonary disease.

## Reach

Our model for health care practitioner engagement with CDSS, which is a proxy for whether health care practitioners were avoiding the intervention by enrolling patients on paper instead of directly on the internet, showed that there was no change in the probability of reporting that a patient was present during enrollment following the beginning of the study (odds ratio [OR] 1.11, 95% CI 0.90-1.35; *P*=.33). However, this model also showed that this probability decreased sharply following the onset of the pandemic (OR 0.46, 95% CI 0.35-0.60; *P*<.001) but then increased (OR per month 1.46, 95% CI 1.38-1.54; *P*<.001). From May 2020 onward, the proportion of enrollments that were conducted directly on the internet was 68% (2518/3701), compared with an average of 58% (46,263/79,505) before the pandemic.

## Effectiveness

Of the participants responding at the 6-month follow-up, 27.3% (552/2020) in the control arm and 29.7% (634/2137) in the intervention arm were abstinent from tobacco at the follow-up. Our pooled estimates of proportions after multiple imputations were 25.9% (95% CI 24.2%-27.6%) and 28.0% (95% CI 26.1%-29.8%), respectively, corresponding to an absolute group difference of 2.1% (95% CI −0.5% to 4.6%). This difference did not meet our threshold for significance ($F_{1,1000.42}$=2.43; *P*=.12). From baseline to the 6-month follow-up, the mean exercise minutes changed from 32 to 113 in the control arm and from 32 to 110 in the intervention arm (group effect: coef=−3.7 minutes, 95% CI −17.8 to 10.4; *P*=.61). The large apparent overall increase in exercise minutes is likely due to regression toward the mean resulting from the use of a cutoff point at baseline. For servings of fruit and vegetables, group means changed from 2.52 at baseline to 2.45 at 6 months in the control group and from 2.64 to 2.42 in the intervention group (incidence rate ratio for intervention group 0.98, 95% CI 0.93-1.02; *P*=.35).

## Adoption

In the intervention group, 1765 participants reported both low levels of fruit and vegetable consumption and low physical activity levels. Of these participants, 1083 (61%) were offered both physical activity and self-monitoring resources for fruit and vegetable consumption. Of the 853 participants who reported low levels of fruit and vegetable consumption (but met the physical activity guidelines) 526 (62%) were offered the self-monitoring resource for diet. Of the 114 participants who reported low physical activity levels (but met the nutrition guidelines) 66 (58%) were offered the self-monitoring resource for physical activity.

## Implementation

Of the 1765 intervention group participants who were offered the appropriate self-monitoring resource for physical activity or fruit and vegetable consumption, 624 (37%) accepted at least one self-monitoring resource.

## Maintenance

The proportion of participants in the intervention group who received an offer of one or both interventions was 67% (932/1402) during the pre–COVID-19 pandemic period; this declined to 60% (791/1329) after the pandemic began in Ontario in March 2020 (difference between these 2 periods: $\chi^2_1$=14.2, *P*<.001).

Monthly proportions during the pandemic period did not differ beyond chance variation ($\chi^2_6$=3.0; *P*=.81), and we did not find evidence of a linear trend (point-biserial correlation=−0.02; *P*=.51). Of the 552 control participants who achieved smoking cessation at 6 months, 438 (79%) responded to a 12-month follow-up. Of these, 322 (74%) were abstinent from smoking at 12 months. In the intervention arm, 634 participants had quit at 6 months. Of these participants, 507 (80%) responded at 12 months and 372 (74%) were not smoking at 12 months.

## Per-Protocol Analysis

In the intervention group, 757 (27.7%) patients received neither a risk communication discussion nor an offer of self-monitoring resource, 810 (29.6%) received one or the other, and 1165 (42.6%) received both. The quit proportions were 27.5% (95% CI 24.0%-31.0%) for those receiving neither resource, 27.6% (95% CI 24.1%-31.0%) for those receiving 1 resource, and 28.5% (95% CI 25.8%-31.3%) for those receiving both resources. The variability in outcomes across the 4 per-protocol levels (including the control arm) was not significant ($F_{3,1289.78}$=0.956; *P*=.41).

### Sensitivity Analyses

Reanalyzing data using only complete cases showed, similar to our main analysis, that the group difference in tobacco abstinence at the 6-month follow-up did not meet our criterion for significance (intervention group: 634/2137, 29.7%; control group: 552/2020, 27.3%; $\chi^2_1$=2.8; *P*=.10). The intervention effect also did not differ for people enrolling after the beginning of the COVID-19 pandemic (test for interaction: *z*=0.58, *P*=.56).

## Discussion

### Principal Findings

The addition of a CDSS for physical activity and fruit and vegetable consumption to a smoking cessation program did not negatively affect 6-month smoking cessation outcomes and did not negatively impact the reach of the smoking cessation program. However, it did not have a significant impact on participants' physical activity or fruit and vegetable consumption at 6 months. That said, we saw that health care practitioners adopted the intervention (offered a self-monitoring resource to eligible participants) with approximately 60% (1083/1675) of their participants. Among the participants who were offered a resource, 37% (624/1675) accepted it. Of the participants who had quit at 6 months and who answered the 12-month survey, 74% (372/507) remained smoke-free, regardless of the study arm.

Given that most of this study took place during the initial phases of COVID-19 pandemic, where many primary care sites partially closed for nonurgent matters at the start of the lockdown and then transitioned to offering virtual services [55], it was important to examine how the pandemic may have impacted the delivery of the intervention. Our analysis showed that the proportion of baseline assessments being completed using the portal decreased sharply at the beginning of the pandemic but increased a couple of months later, exceeding the prepandemic proportion. Our analysis also showed that, compared with before the pandemic, a smaller proportion of participants received a physical activity or diet intervention.

### Interpretation and Comparison With Prior Work

Taken together, these results indicate that we can modify a smoking cessation program to be more holistic without negatively impacting smoking cessation, the single most important behavior change for reducing chronic disease-related mortality [52]. While there is ample evidence that modifying multiple health behaviors improves population health and reduces health care expenditures [56,57], there is insufficient research on effective ways to implement these changes. Furthermore, when multiple health behavior changes are necessary, knowing the impact of changing one behavior on another is important for health care practitioners as well as for decision-makers. For example, some research findings show that when physical activity and diet interventions were added to smoking cessation programs, there is a reduction in smoking [58,59]. However, other studies have reported either an adverse effect or no effect when physical activity and diet was integrated into a smoking cessation programming [60].

The results also showed that health care practitioners adopted the intervention (offered a brief intervention or self-monitoring resource) with approximately 60% (1083/1765) of their patients who were eligible, and 37% (624/1765) of these patients accepted the resource. The adoption outcome for this intervention was considerably higher than what we have seen for alcohol use in the STOP program, 21% of patients who drank alcohol above guidelines were offered an educational resource [33], and elsewhere for other health behavior interventions [61]. The implementation outcome (acceptance of resource by patient) in this trial was also higher than the ones reported for delivering a mood management intervention [61,62].

### COVID-19 Pandemic

The results of this study were likely affected by the onset of the COVID-19 pandemic. As we mentioned earlier, most of this study took place during the COVID-19 pandemic, which affected the context in which this study occurred, including the outer setting, the inner setting, and the health care practitioners and the participants' behavior. From previous studies, we know that there were significantly fewer new enrollments and subsequent visits to the STOP program during the initial phase of the pandemic and that there was an increased number of STOP participants who reported being unemployed, as well as having substance use, mental health, and physical health diagnoses [55]. We also know that in the STOP program, the likelihood of successful smoking cessation after treatment dropped during the pandemic [63]. All of these factors might have affected both the implementation and effectiveness outcomes. For example, practitioners might have been less likely to recommend eating more fruit and vegetables if a participant reported that they were recently unemployed but more likely to recommend physical activity, which can be done without incurring a financial cost. Practitioners were also impacted by greater time constraints related to COVID-19 pandemic (ie, redeployment) and may not have been able to spend as much time addressing physical activity and diet with their participants [45].

While there was a large drop in the use of the portal to complete STOP enrollments during the SOE, within a couple of months the proportion of enrollments using the portal exceeded the prepandemic proportions (68%, compared with an average of 58% before the pandemic). This may be indicative of the positive impact the virtual care can have on a patient's treatment experience, including minimizing logistical burdens around appointment management [64]. Since the appointments can be conducted over the phone, practitioners may have found it convenient to do the enrollment directly on the portal, instead of having the patient complete it on paper.

### Strengths and Limitations

This study had several strengths: it tested a simple web-based CDSS to facilitate the delivery of an intervention to address physical activity and diet among people who smoke. The results are generalizable given that we tested it in a large geographical area, with sites offering different types of primary care with varying staffing models and patient populations (family health teams, community health centers, and nurse practitioner–led clinics). By conducting a hybrid effectiveness-implementation trial, we showed that the addition of the CDSS did not negatively affect the reach of the program or the effectiveness of smoking cessation and that it was adopted by health care practitioners. However, the results also show that there is room for improvement with respect to implementation increasing the likelihood of practitioners offering self-monitoring resources to their patients.

There are a few limitations that need to be acknowledged. Our primary and secondary outcomes use self-reported measures, which may not accurately represent behavior changes in patients. Historical self-reported outcomes rely on patients accurately remembering their current behavior; thus, there is an increased risk of error and bias. However, these self-reported measures have been validated, and there is no reason to believe that any self-report bias would differ between the intervention and control groups. This study excluded French-speaking participants. We might have reduced the representativeness of our sample by excluding French-speaking participants; however, French-speaking participants make up a very small proportion of the STOP program, and we do not believe that their responses to the intervention would have differed from the included patients.

The fruit and vegetable question includes 100% fruit juice as part of the servings, which is not in line with the latest Canada’s Food Guideline [65], which was published after this study was created. However, we were only looking for overall changes in fruit and vegetable consumption as an indicator that could be used in primary care and differences between the intervention and control groups. In terms of limitations at the practitioner level, we have no way of verifying whether the practitioners in the intervention group actually acted on the CDSS recommendations and guidance as intended. Furthermore, while patients were blinded to their treatment allocation, health care practitioners were not, as we used patient-level randomization. Health care practitioners treating the control group participants could still address physical activity and diet with these participants. This lack of health care practitioner blinding increases the risk of group contamination as the health care practitioner could take their learnings from the intervention group patients and apply it to the control group participants. However, the physical activity and fruit and vegetable consumption self-monitoring resources in the CDSS were only available to participants in the intervention group, which minimized the risk of any cross-contamination. Thus, future studies should consider including additional process measures to accurately track any potential cross-contamination in the control group. This can be as simple as routinely asking control group practitioners whether they provided their own intervention to the patient.

## Conclusions

The introduction of a CDSS that guides health care practitioners to address multiple health behaviors among their patients does not seem to affect smoking cessation success. Although additional research is needed, these findings demonstrate that the CDSS can be used to introduce holistic treatment approaches within a primary care smoking cessation program. Thus, the CDSS could be a potential solution to break with the siloed approach of behavior change, which has dominated many care treatment approaches.

## Acknowledgments

The authors thank the staff and patients who participated in this study. PS would like to acknowledge the salary support for his clinician-scientist position from the Centre for Addiction and Mental Health and the Department of Family and Community Medicine at the University of Toronto. This research was supported by the Public Health Agency of Canada (grant 1617-HQ-000045) and the Medical Psychiatry Alliance.

## Authors' Contributions

All authors have contributed to the design, execution, and writing of this manuscript and approved the submitted version.

## Conflicts of Interest

The authors declare that they have no conflicts of interest with respect to this manuscript; however, some authors have general disclosures to report. PS reports receiving grants or salary or research support from the Centre for Addiction and Mental Health, Health Canada, Ontario Ministry of Health, Canadian Institutes of Health Research, Public Health Agency of Canada, Medical Psychiatry Alliance, Canadian Cancer Society Research Institute, Cancer Care Ontario, and Ontario Institute for Cancer Research. PS also reports receiving funding or honoraria from the following commercial organizations: Pfizer Inc Canada, Bhasin Consulting Fund Inc, and Patient-Centered Outcomes Research Institute. PS reports receiving consulting fees from Pfizer Inc Canada, Johnson & Johnson Group of Companies, and Myelin and Associates. Through an open tender process, Johnson & Johnson, Novartis, and Pfizer Inc are vendors of record for providing smoking cessation pharmacotherapy, free or discounted, for research studies in which PS is the principal investigator or coinvestigator. OCM reports receiving grant or salary support from the Centre for Addiction and Mental Health and AMS Healthcare. MH reports receiving consulting fees from Alkermes. WKD reports grant funding from Pfizer Inc and the Medical Psychiatry Alliance. WKD also reports stock ownership in Abbott Laboratories.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1207 KB - jmir_v24i9e37900_app1.pdf ]

---

## References

1. Manuel DG, Perez R, Bennett C, Rosella L, Taljaard M, Roberts M, et al. Seven more years: the impact of smoking, alcohol, diet, physical activity and stress on health and life expectancy in Ontario. Institute for Clinical Evaluative Science. 2012. URL: https://www.ices.on.ca/flip-publication/seven-years/files/assets/basic-html/index.html#1 [accessed 2022-01-30]
2. Ng R, Sutradhar R, Yao Z, Wodchis WP, Rosella LC. Smoking, drinking, diet and physical activity-modifiable lifestyle risk factors and their associations with age to first chronic disease. Int J Epidemiol 2020 Feb 01;49(1):113-130 [FREE Full text] [doi: 10.1093/ije/dyz078] [Medline: 31329872]
3. Li Y, Pan A, Wang DD, Liu X, Dhana K, Franco OH, et al. Impact of healthy lifestyle factors on life expectancies in the US population. Circulation 2018 Jul 24;138(4):345-355 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.032047] [Medline: 29712712]
4. Manuel DG, Perez R, Sanmartin C, Taljaard M, Hennessy D, Wilson K, et al. Measuring burden of unhealthy behaviours using a multivariable predictive approach: life expectancy lost in Canada attributable to smoking, alcohol, physical inactivity, and diet. PLoS Med 2016 Aug 16;13(8):e1002082 [FREE Full text] [doi: 10.1371/journal.pmed.1002082] [Medline: 27529741]
5. deRuiter WK, Cairney J, Leatherdale ST, Faulkner GE. A longitudinal examination of the interrelationship of multiple health behaviors. Am J Prev Med 2014 Sep;47(3):283-289. [doi: 10.1016/j.amepre.2014.04.019] [Medline: 25145617]
6. Meader N, King K, Moe-Byrne T, Wright K, Graham H, Petticrew M, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. BMC Public Health 2016 Jul 29;16:657 [FREE Full text] [doi: 10.1186/s12889-016-3373-6] [Medline: 27473458]
7. Leatherdale ST. An examination of the co-occurrence of modifiable risk factors associated with chronic disease among youth in the COMPASS study. Cancer Causes Control 2015 Apr;26(4):519-528. [doi: 10.1007/s10552-015-0529-0] [Medline: 25673505]
8. Minian N, Corrin T, Lingam M, deRuiter WK, Rodak T, Taylor VH, et al. Identifying contexts and mechanisms in multiple behavior change interventions affecting smoking cessation success: a rapid realist review. BMC Public Health 2020 Jun 12;20(1):918 [FREE Full text] [doi: 10.1186/s12889-020-08973-2] [Medline: 32532233]
9. Butterfield RM, Park ER, Puleo E, Mertens A, Gritz ER, Li FP, et al. Multiple risk behaviors among smokers in the childhood cancer survivors study cohort. Psychooncology 2004 Sep;13(9):619-629. [doi: 10.1002/pon.764] [Medline: 15334530]
10. Cecchini M, Sassi F, Lauer JA, Lee YY, Guajardo-Barron V, Chisholm D. Tackling of unhealthy diets, physical inactivity, and obesity: health effects and cost-effectiveness. Lancet 2010 Nov 20;376(9754):1775-1784. [doi: 10.1016/S0140-6736(10)61514-0] [Medline: 21074255]
11. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
12. Kahler CW, Metrik J, Spillane NS, Day A, Leventhal AM, McKee SA, et al. Acute effects of low and high dose alcohol on smoking lapse behavior in a laboratory analogue task. Psychopharmacology (Berl) 2014 Dec;231(24):4649-4657 [FREE Full text] [doi: 10.1007/s00213-014-3613-3] [Medline: 24858377]
13. Parrott AC. Does cigarette smoking cause stress? Am Psychol 1999 Oct;54(10):817-820. [doi: 10.1037//0003-066x.54.10.817] [Medline: 10540594]
14. Taylor AH, Ussher MH, Faulkner G. The acute effects of exercise on cigarette cravings, withdrawal symptoms, affect and smoking behaviour: a systematic review. Addiction 2007 Apr;102(4):534-543. [doi: 10.1111/j.1360-0443.2006.01739.x] [Medline: 17286639]
15. Deruiter WK, Faulkner G, Cairney J, Veldhuizen S. Characteristics of physically active smokers and implications for harm reduction. Am J Public Health 2008 May;98(5):925-931. [doi: 10.2105/AJPH.2007.120469] [Medline: 18381990]
16. Haasova M, Warren FC, Ussher M, Janse Van Rensburg K, Faulkner G, Cropley M, et al. The acute effects of physical activity on cigarette cravings: systematic review and meta-analysis with individual participant data. Addiction 2013 Jan;108(1):26-37. [doi: 10.1111/j.1360-0443.2012.04034.x] [Medline: 22861822]
17. Roberts V, Maddison R, Simpson C, Bullen C, Prapavessis H. The acute effects of exercise on cigarette cravings, withdrawal symptoms, affect, and smoking behaviour: systematic review update and meta-analysis. Psychopharmacology (Berl) 2012 Jul;222(1):1-15. [doi: 10.1007/s00213-012-2731-z] [Medline: 22585034]
18. Pánková A, Králíková E, Štepánková L, Zvolska K, Bortlícek Z, Bláha M, et al. Weight concerns associated with delay in quit date but not treatment outcomes: a Czech republic experience. Nicotine Tob Res 2017 Dec 13;20(1):89-94. [doi: 10.1093/ntr/ntw276] [Medline: 27729514]
19. Lancaster T, Stead L, Silagy C, Sowden A. Effectiveness of interventions to help people stop smoking: findings from the Cochrane Library. BMJ 2000 Aug 05;321(7257):355-358 [FREE Full text] [doi: 10.1136/bmj.321.7257.355] [Medline: 10926597]

20. Daniel JZ, Cropley M, Fife-Schaw C. Acute exercise effects on smoking withdrawal symptoms and desire to smoke are not related to expectation. Psychopharmacology (Berl) 2007 Nov;195(1):125-129. [doi: 10.1007/s00213-007-0889-6] [Medline: 17653531]
21. Crowley RA, Kirschner N, Health and Public Policy Committee of the American College of Physicians. The integration of care for mental health, substance abuse, and other behavioral health conditions into primary care: executive summary of an American College of Physicians position paper. Ann Intern Med 2015 Aug 18;163(4):298-299 [FREE Full text] [doi: 10.7326/M15-0510] [Medline: 26121401]
22. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998 Oct 21;280(15):1339-1346. [doi: 10.1001/jama.280.15.1339] [Medline: 9794315]
23. Schiff GD, Rucker TD. Computerized prescribing: building the electronic infrastructure for better medication usage. JAMA 1998 Apr 01;279(13):1024-1029. [doi: 10.1001/jama.279.13.1024] [Medline: 9533503]
24. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020 Feb 6;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
25. Balas EA, Weingarten S, Garb CT, Blumenthal D, Boren SA, Brown GD. Improving preventive care by prompting physicians. Arch Intern Med 2000 Feb 14;160(3):301-308. [doi: 10.1001/archinte.160.3.301] [Medline: 10668831]
26. Bennett JW, Glasziou PP. Computerised reminders and feedback in medication management: a systematic review of randomised controlled trials. Med J Aust 2003 Mar 03;178(5):217-222. [doi: 10.5694/j.1326-5377.2003.tb05166.x] [Medline: 12603185]
27. 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]
28. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. Arch Intern Med 2003 Jun 23;163(12):1409-1416. [doi: 10.1001/archinte.163.12.1409] [Medline: 12824090]
29. Durieux P, Trinquart L, Colombet I, Niès J, Walton R, Rajeswaran A, et al. Computerized advice on drug dosage to improve prescribing practice. Cochrane Database Syst Rev 2008 Jul 16(3):CD002894. [doi: 10.1002/14651858.CD002894.pub2] [Medline: 18646085]
30. Meisner JK, Yu S, Lowery R, Liang W, Schumacher KR, Burrows HL. Clinical decision support tool for elevated pediatric blood pressures. Clin Pediatr (Phila) 2022 Jun;61(5-6):428-439. [doi: 10.1177/00099228221087804] [Medline: 35383471]
31. Huryk LA. Information systems and decision support systems. Am J Nurs 2012 Jan;112(1):62-65. [doi: 10.1097/01.NAJ.0000410365.51061.81] [Medline: 22222406]
32. Birtwhistle R, Williamson T. Primary care electronic medical records: a new data source for research in Canada. CMAJ 2015 Mar 03;187(4):239-240 [FREE Full text] [doi: 10.1503/cmaj.140473] [Medline: 25421989]
33. Minian N, Baliunas D, Noormohamed A, Zawertailo L, Giesbrecht N, Hendershot CS, et al. The effect of a clinical decision support system on prompting an intervention for risky alcohol use in a primary care smoking cessation program: a cluster randomized trial. Implement Sci 2019 Aug 23;14(1):85 [FREE Full text] [doi: 10.1186/s13012-019-0935-x] [Medline: 31443663]
34. Canadian Physical Activity Guidelines for Adults 18-64 years. Canadian Society for Exercise Physiology. 2011. URL: https://csepguidelines.ca/adults-18-64/ [accessed 2021-08-01]
35. Eating Well with Canada's Food Guide 2007. Government of Canada. 2007. URL: https://www.canada.ca/en/health-canada/services/canada-food-guide/about/history-food-guide/eating-well-with-canada-food-guide-2007.html [accessed 2021-08-01]
36. Panigutti C, Perotti A, Panisson A, Bajardi P, Pedreschi D. FairLens: auditing black-box clinical decision support systems. Inf Process Manag 2021 Sep;58(5):102657 [FREE Full text] [doi: 10.1016/j.ipm.2021.102657]
37. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
38. Minian N, Lingam M, Moineddin R, Thorpe KE, Veldhuizen S, Dragonetti R, et al. Impact of a Web-based clinical decision support system to assist practitioners in addressing physical activity and/or healthy eating for smoking cessation treatment: protocol for a hybrid type I randomized controlled trial. JMIR Res Protoc 2020 Sep 29;9(9):e19157 [FREE Full text] [doi: 10.2196/19157] [Medline: 32990250]
39. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
40. Minian N, Noormohamed A, Lingam M, Zawertailo L, Le Foll B, Rehm J, et al. Integrating a brief alcohol intervention with tobacco addiction treatment in primary care: qualitative study of health care practitioner perceptions. Addict Sci Clin Pract 2021 Mar 16;16(1):17 [FREE Full text] [doi: 10.1186/s13722-021-00225-x] [Medline: 33726843]

41. Wandersman A, Duffy J, Flaspohler P, Noonan R, Lubell K, Stillman L, et al. Bridging the gap between prevention research and practice: the interactive systems framework for dissemination and implementation. Am J Community Psychol 2008 Jun;41(3-4):171-181. [doi: 10.1007/s10464-008-9174-z] [Medline: 18302018]
42. Minian N, Lingam M, deRuiter WK, Dragonetti R, Selby P. Co-designing behavior change resources with treatment-seeking smokers: engagement events' findings. Front Public Health 2021 Mar 15;9:555449 [FREE Full text] [doi: 10.3389/fpubh.2021.555449] [Medline: 33791263]
43. Selby P. Addressing Modifiable Risk Factors to Improve Smoking Cessation Outcomes in the STOP Program. Nicotine Dependence Service- CAMH. 2019. URL: https://www.youtube.com/watch?v=TIsG2V13JT8&list=PLmLKlp1R6077gOXIyxwToXAZbVrG7GnZx&index=4&ab_channel=teachproject [accessed 2022-03-10]
44. van Achterberg T, Huisman-de Waal GG, Ketelaar NA, Oostendorp RA, Jacobs JE, Wollersheim HC. How to promote healthy behaviours in patients? An overview of evidence for behaviour change techniques. Health Promot Int 2011 Jun;26(2):148-162 [FREE Full text] [doi: 10.1093/heapro/daq050] [Medline: 20739325]
45. Glazier RH, Green ME, Wu FC, Frymire E, Kopp A, Kiran T. Shifts in office and virtual primary care during the early COVID-19 pandemic in Ontario, Canada. CMAJ 2021 Feb 08;193(6):E200-E210 [FREE Full text] [doi: 10.1503/cmaj.202303] [Medline: 33558406]
46. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. Am J Public Health 1994 Jul;84(7):1086-1093. [doi: 10.2105/ajph.84.7.1086] [Medline: 8017530]
47. Studts JL, Ghate SR, Gill JL, Studts CR, Barnes CN, LaJoie AS, et al. Validity of self-reported smoking status among participants in a lung cancer screening trial. Cancer Epidemiol Biomarkers Prev 2006 Oct;15(10):1825-1828. [doi: 10.1158/1055-9965.EPI-06-0393] [Medline: 17035388]
48. Vartiainen E, Seppälä T, Lillsunde P, Puska P. Validation of self reported smoking by serum cotinine measurement in a community-based study. J Epidemiol Community Health 2002 Mar;56(3):167-170 [FREE Full text] [doi: 10.1136/jech.56.3.167] [Medline: 11854334]
49. Wong SL, Shields M, Leatherdale S, Malaison E, Hammond D. Assessment of validity of self-reported smoking status. Health Rep 2012 Mar;23(1):47-53 [FREE Full text] [Medline: 22590805]
50. Liu I, Moy M, Estrada E, Rippberger E, Nguyen H. An "exercise vital sign" is a valid proxy measure of physical activity in COPD in routine clinical care. Trans J Am Coll Sports Med 2017 Dec;2(23):148-152.
51. Di Leo G, Sardanelli F. Statistical significance: P value, 0.05 threshold, and applications to radiomics-reasons for a conservative approach. Eur Radiol Exp 2020 Mar 11;4(1):18 [FREE Full text] [doi: 10.1186/s41747-020-0145-y] [Medline: 32157489]
52. Feenstra TL, Van Baal P, Hoogenveen R, Vijgen SM, Bemelmans WJ. The cost-effectiveness of smoking cessation interventions: accounting for medical costs in longer life expectancies. Value Health 2006 Nov;9(6):A207 [FREE Full text] [doi: 10.1016/s1098-3015(10)63218-4]
53. Lumley T. mitools: Tools for Multiple Imputation of Missing Data. The Comprehensive R Archive Network. 2019 Apr 26. URL: https://cran.r-project.org/web/packages/mitools/index.html [accessed 2022-02-03]
54. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. URL: https://www.r-project.org/ [accessed 2022-02-03]
55. Minian N, Veldhuizen S, Tanzini E, Duench S, deRuiter WK, Barker M, et al. Changes in the reach of a smoking cessation program in Ontario, Canada, during the COVID-19 pandemic: a cross-sectional study. CMAJ Open 2021 Oct 19;9(4):E957-E965 [FREE Full text] [doi: 10.9778/cmajo.20210053] [Medline: 34667076]
56. Jepson RG, Harris FM, Platt S, Tannahill C. The effectiveness of interventions to change six health behaviours: a review of reviews. BMC Public Health 2010 Sep 08;10:538 [FREE Full text] [doi: 10.1186/1471-2458-10-538] [Medline: 20825660]
57. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. World Health Organization. 2013. URL: https://apps.who.int/iris/rest/bitstreams/442296/retrieve [accessed 2022-01-25]
58. Marcus BH, Albrecht AE, King TK, Parisi AF, Pinto BM, Roberts M, et al. The efficacy of exercise as an aid for smoking cessation in women: a randomized controlled trial. Arch Intern Med 1999 Jun 14;159(11):1229-1234. [doi: 10.1001/archinte.159.11.1229] [Medline: 10371231]
59. Perkins KA, Marcus MD, Levine MD, D'Amico D, Miller A, Broge M, et al. Cognitive-behavioral therapy to reduce weight concerns improves smoking cessation outcome in weight-concerned women. J Consult Clin Psychol 2001 Aug;69(4):604-613. [Medline: 11550727]
60. Hall SM, Tunstall CD, Vila KL, Duffy J. Weight gain prevention and smoking cessation: cautionary findings. Am J Public Health 1992 Jun;82(6):799-803. [doi: 10.2105/ajph.82.6.799] [Medline: 1585959]
61. Hargraves D, White C, Frederick R, Cinibulk M, Peters M, Young A, et al. Implementing SBIRT (Screening, Brief Intervention and Referral to Treatment) in primary care: lessons learned from a multi-practice evaluation portfolio. Public Health Rev 2017 Dec 29;38:31 [FREE Full text] [doi: 10.1186/s40985-017-0077-0] [Medline: 29450101]
62. Minian N, Ahad S, Ivanova A, Veldhuizen S, Zawertailo L, Ravindran A, et al. The effectiveness of generic emails versus a remote knowledge broker to integrate mood management into a smoking cessation programme in team-based primary

care: a cluster randomised trial. Implement Sci 2021 Mar 20;16(1):30 [FREE Full text] [doi: 10.1186/s13012-021-01091-6] [Medline: 33743777]

63. Veldhuizen S, Selby P, Wong B, Zawertailo L. Effect of COVID-19 on smoking cessation outcomes in a large primary care treatment programme: an observational study. BMJ Open 2021 Aug 26;11(8):e053075 [FREE Full text] [doi: 10.1136/bmjopen-2021-053075] [Medline: 34446503]
64. Kelley LT, Phung M, Stamenova V, Fujioka J, Agarwal P, Onabajo N, et al. Exploring how virtual primary care visits affect patient burden of treatment. Int J Med Inform 2020 Sep;141:104228. [doi: 10.1016/j.ijmedinf.2020.104228] [Medline: 32683311]
65. Canada’s Dietary Guidelines. Health Canada, Government of Canada. 2019. URL: https://food-guide.canada.ca/en/guidelines/ [accessed 2022-01-20]

## Abbreviations

**CDSS:** clinical decision support system
**ISF:** Interactive Systems Framework
**OR:** odds ratio
**RE-AIM:** Reach, Effectiveness, Adoption, Implementation, and Maintenance
**SOE:** state of emergency
**STOP:** Smoking Treatment for Ontario Patients

---

*Edited by R Kukafka, G Eysenbach; submitted 11.03.22; peer-reviewed by K Uludag, A Videira-Silva, FM Calisto, H Veldandi; comments to author 11.05.22; revised version received 29.06.22; accepted 29.07.22; published 30.09.22.*

*Please cite as:*
*Minian N, Lingam M, Moineddin R, Thorpe KE, Veldhuizen S, Dragonetti R, Zawertailo L, Taylor VH, Hahn M, deRuiter WK, Melamed OC, Selby P*
*The Impact of a Clinical Decision Support System for Addressing Physical Activity and Healthy Eating During Smoking Cessation Treatment: Hybrid Type I Randomized Controlled Trial*
*J Med Internet Res 2022;24(9):e37900*
*URL: https://www.jmir.org/2022/9/e37900*
*doi:10.2196/37900*
*PMID:36178716*

Original Paper

# Information and Communication Technologies in Patients With Immune-Mediated Inflammatory Diseases: Cross-sectional Survey

Esther Chamorro-de-Vega[1,2], PharmD, PhD; Rosa Romero-Jiménez[2], PharmD, PhD; Vicente Escudero-Vilaplana[2], PharmD, PhD; Arantza Ais-Larisgoitia[2], PharmD; María Elena Lobato Matilla[2], PharmD; Carlos M González[3], MD, PhD; Luis Menchén[4,5], MD, PhD; Ofelia Baniandrés[6], MD, PhD; Lucía Ibares-Frias[7], MD, PhD; Carmen Lobo-Rodríguez[8], PhD; Ana Herranz-Alonso[2], PharmD, PhD; María Sanjurjo[2], PharmD

[1]Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[2]Department of Pharmacy, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[3]Department of Rheumatology, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[4]Department of Gastroenterology, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[5]Departamento de Medicina, Facultad de Medicina, Universidad Complutense, Madrid, Spain
[6]Department of Dermatology, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[7]Department of Ophthalmology, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[8]Nursing Group, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain

**Corresponding Author:**
Esther Chamorro-de-Vega, PharmD, PhD
Department of Pharmacy
Hospital General Universitario Gregorio Marañón
Instituto de Investigación Sanitaria Gregorio Marañón
Dr. Esquerdo 46
Madrid, 28007
Spain
Phone: 34 915 867 714
Email: echamorro@salud.madrid.org

## Abstract

**Background:** Information and communication technologies (ICTs) are changing the traditional health care model and redefining personalized health. ICTs offer effective communication and real-time monitoring of patients and provide additional data to support clinical decision-making, improve the quality of care, and contribute to the empowerment of patients. However, evidence on the use of ICTs and digital preferences of immune-mediated inflammatory disease (IMID) patients is scarce.

**Objective:** The aim of this study is to describe the degree of use of ICTs in patients with IMIDs (including rheumatic diseases, inflammatory bowel diseases, and psoriasis), identify their needs, and analyze their interest in the use of apps as tools for better management of their disease.

**Methods:** A questionnaire was created by a multidisciplinary team including pharmacists, rheumatologists, gastroenterologists, dermatologists, and nurses with experience in ICTs applied to the field of IMID. The survey included 27 questions organized into 3 blocks: (1) sociodemographic characteristics, (2) ICT use for health-related information, and (3) patient expectations about mobile health.

**Results:** A total of 472 questionnaires were analyzed. Overall, 52.9% (250/472) of patients were diagnosed with a rheumatologic disease, 39.4% (186/472) with inflammatory bowel disease, and 12.3% (58/472) with psoriasis. The state of health was considered good by 45.6% (215/472) of patients. Patients were interested in staying informed about health issues in 86.9% (410/427) of cases and sought health-related information mainly from the internet (334/472, 70.8%) and health care professionals (318/472, 67.4%). Overall, 13.6% (64/472) did not trust the health information they found in internet. Of the patients, 42.8% (202/472) had a health app, and 42.2% (199/472) had found it on their own. Patients would like a health app to help mainly to manage appointments (281/472, 59.5%), obtain information about their diseases and treatments (274/472, 58.1%), and get in contact with health professionals (250/472, 53.0%). Overall, 90.0% (425/472) of patients reported they would use an app to manage their IMID if their health professional recommended it, and 58.0% (274/472) would pay or probably be willing to pay for it.

**Conclusions:** IMID patients were very interested in finding health-related information via ICTs, especially using smartphones and apps recommended by health professionals. Appointment management, advice on disease and treatment management, and personalized communication with health professionals were the most desired app features identified. Health professionals should play an essential role in recommending and validating these tools to ensure they are of high quality.

***(J Med Internet Res 2022;24(9):e37445)*** doi:10.2196/37445

**KEYWORDS**

mHealth; app; information and communication technologies; immune-mediated inflammatory diseases; IMID; cross-sectional survey; survey; monitoring; clinical support; clinicians; quality of care; patient care; mobile app; tool; management tool

# Introduction

Immune-mediated inflammatory diseases (IMID) are a large group of chronic diseases of high prevalence and a high socio-health impact [1]. Patients with IMID are generally middle age, with diverse clinical manifestations, and reduced quality of life what makes its handling complex [2,3]. These patients often report a lack of understanding of their disease, demand information in real time, and sometimes require unscheduled health care due to clinical decompensations, causing a high consumption of health care resources [4-6]. This situation represents a challenge for the usual structure of health care services, making it necessary to develop new models of clinical management that allow making intensive and regular disease monitoring as well as maintaining the sustainability of the system.

The recent emergence of information technologies (ICTs), concretely mobile health (mHealth), has been positioned as key tools to improve the quality of health care since it helps both professionals and patients. On the one hand, ICTs allow health care professionals to access the patient's clinical information in real time, improving risk monitoring and enabling proactive treatment. On the other hand, these tools facilitate connected care and allow patients to have direct communication with their health professionals [7-11]. Moreover, ICTs favored the conciliation of patients with other activities of their daily life, promoting their autonomy, and streamlining processes and bureaucratic procedures. Thus, ICTs, especially apps, have the potential for promoting self-management, improving standard clinical care, and reducing the impact of the growing burden of IMID on health care resource utilization [12-15].

There is a wide variety of apps for IMID patients on the market with numerous features and functionalities but a lack of validity and scientific reliability [12,14,16]. The lack of regulation of these tools makes health care professionals primarily responsible for ensuring the validity of the tools they recommend to their patients. However, information and evidence about the real use of ICTs in IMID patients and the needs and expectations patients have for these technologies are scarce. This is essential for further validation and improvement of health-related app, from both a clinical and patient perspective [17-19]. With this study, we aimed to describe the degree of use of information and communication technologies in patients with IMID, identify their needs, and analyze their interest in the use of apps as a tool for better management of their disease.

# Methods

## Study Design and Setting

We performed a descriptive study in a tertiary care teaching hospital of the Madrid Public Health Service (Spain). This hospital has a coverage of the entire portfolio of services for a population of 325,000 inhabitants distributed in 11 basic health areas. This hospital has a comprehensive care center, a reference center for the management of IMID care, for patients with IMID.

## Study Patients

Patients diagnosed with IMID, including rheumatic disorders, inflammatory bowel disease, and psoriasis and treated with biological or targeted therapies, were eligible for the study. Patients under 18 years, those treated in a clinical trial, and those who had trouble understanding the questions due to language or cultural barriers were excluded. All patients provided written informed consent and agreed to participate.

## Data Collection

Data were collected from December 2020 to July 2021 during a single visit through an anonymous patient self-questionnaire. To design the questionnaire, the survey created by Spanish National Observatory of Telecommunications and Information Society [20] was taken as a model and modified and adapted to our population by a multidisciplinary team including pharmacists, rheumatologists, gastroenterologists, dermatologists, and nurses with experience in ICTs applied to the field of IMID. Finally, a 27-item questionnaire was designed (Multimedia Appendix 1). The survey comprised 3 main blocks. Block I: patient sociodemographic characteristics (questions 1-6); block II: ICT use for health-related information (questions 7-17); and block III: patient expectations about mHealth (questions 18-27). With respect to the types of questions, the questionnaire comprised dichotomous statements (2, 10, 15-18, 20, 25, and 26) with true or false response options or polytomous statements (3-9, 11-14, 19, and 21-24) categorical variables. Some statements were multiple choice (5, 7-9, 11-13, 21-24). Ten patients completed a draft paper questionnaire to validate the fact that it was well understood. Regarding the question of whether patients would be willing to pay to download an app, the cost of the app was taken from the review published by Collado-Borrell et al [21] who showed an average price of apps for patients with cancer of 2.15€ (US $2.15).

A paper copy of the questionnaire was handed to every patient in the infusion center, in pharmaceutical consultation, or in the waiting rooms of the comprehensive care center. Alternatively, patients could take the questionnaire home and return the completed form at their next appointment.

### Statistical Analysis

Data were analyzed using SPSS for Windows (version 21.0, IBM Corp) software. Variables were analyzed using descriptive statistics. Normality was analyzed by means of the Kolmogorov-Smirnov test. Numeric variables were compared with the Student *t* test or the Mann-Whitney test. The association between qualitative variables was studied using the Pearson chi-square test or Fisher exact test. The corresponding measures of association and risk were calculated along with their confidence intervals. Results with a value of *P*<.05 were considered statistically significant.

### Ethics Approval

The study protocol was approved by the Ethics Committee of the Hospital General Universitario Gregorio Marañón (approval number: IMID-HGUGM.01), in accordance with the principles of the Declaration of Helsinki. The questionnaire did not include any information about the personal data of patients to ensure data confidentiality.

## Results

### Patient Sociodemographic Characteristics

Of the questionnaires distributed, there was a participation of 95.2% (472/496). The remaining 24 questionnaires were not analyzed because they were not understandable. Patient sociodemographic characteristics are shown in Table 1. The median age of patients was 50.5 years (IQR 40.0-59.8). Of all participants, 53.2% (251/472) were women and 44.7% (211/472) had a university education.

Overall, 52.9% (250/472) of patients were diagnosed with a rheumatologic disease, 21.8% (103/472) rheumatoid arthritis, 16.7% (79/472) spondylarthritis, 13.1% (62/472) psoriatic arthritis, and 1.3% (6/472) other. Among nonrheumatologic patients, 39.4% (186/472) were inflammatory bowel disease, 71.5% (133/186) were Crohn disease, 28.5% (53/186) were ulcerative colitis, and 12.3% (58/472) were psoriasis. Overall, 12.1% (57/472) of patients were diagnosed with more than one concomitant IMID. Over half (269/472, 57.0%) of patients considered their state of health was either good 45.6% (215/472) or very good 11.4% (54/472).

**Table 1.** Patient sociodemographic characteristics (n=472).

| Demographic characteristics | Value |
|---|---|
| **Sex, n (%)** | |
| Female | 251 (53.2) |
| Male | 221 (46.8) |
| **Age (years), mean (SD)** | 49.9 (13.4) |
| ≤35, n (%) | 74 (15.7) |
| 36-50, n (%) | 159 (33.7) |
| 51-60, n (%) | 128 (27.1) |
| ≥61, n (%) | 106 (22.4) |
| No response, n (%) | 5 (1.1) |
| **Current occupation, n (%)** | |
| Works | 280 (59.3) |
| Retired | 88 (18.6) |
| Unemployed | 53 (11.2) |
| Homemaker | 35 (7.4) |
| Study | 15 (3.2) |
| No response | 1 (0.2) |
| **Level of education, n (%)** | |
| No education or incomplete primary education | 10 (2.1) |
| Primary education | 80 (16.9) |
| Secondary education | 167 (35.4) |
| University education | 211 (44.7) |
| No response | 4 (0.8) |
| **Type of IMID[a,b], n (%)** | |
| Crohn disease | 133 (28.2) |
| Rheumatoid arthritis | 103 (21.8) |
| Spondyloarthritis | 79 (16.7) |
| Psoriatic arthritis | 62 (13.1) |
| Psoriasis | 58 (12.3) |
| Ulcerative colitis | 53 (11.2) |
| Other | 48 (10.2) |
| More than one IMID | 57 (12.1) |
| **How would you rate your overall health? n (%)** | |
| Very good | 54 (11.4) |
| Good | 215 (45.6) |
| Average | 172 (36.4) |
| Poor | 27 (5.7) |
| Very poor | 1 (0.2) |
| No response | 3 (0.6) |

[a] Multiple choice question.

[b] IMID: immune-mediated inflammatory disease.

### ICT Use for Health-Related Information

Patients were questioned concerning internet use to address their habits regarding information search (Table 2). Most of the patients had heard the term app and smartphone (359/472, 76.1%, and 347/472, 73.5, respectively) before, but only 17.6% (83/472) had heard the term wearable. Mobile phones and desktops were the devices most used for searching for information on the internet (433/472, 91.7%, and 308/472, 65.3%, respectively). The mobile phone was used daily to search for information on the internet by 65.5% (309/472) of patients. Figure 1 shows the frequency with which patients used different ICTs to search for information on the internet.

Patients sought health-related information mainly from the internet and health care professionals (334/472, 70.8%, and 318/472, 67.4%, respectively). Of patients who consulted the internet, 73.5% (347/472) searched through Google and 33.3% (157/472) through social networks (YouTube, Twitter, and Facebook, among others) and 24.2% (114/472) used medical societies. About half of patients looked for health information to obtain information about disease prevention, healthy lifestyles, and health care (239/472, 50.6%) and to find information about the treatment their doctor had prescribed for them (215/472, 45.6%).

We observed statistically significant differences depending on the level of education, age, sex, and IMID type. First, patients with secondary or university education and younger patients searched for more information on health in the internet (*P*<.001). Patients with secondary or university education sought more information about health centers or health professionals (*P*<.001), regarding symptoms and learning about potential diseases (*P*=.003), and regarding their treatment (*P*=.02). Men sought more information regarding symptoms and learning about potential diseases (*P*=.02), while women wanted to get in contact with other people with similar health problems (*P*=.004). Patients with ulcerative colitis were most likely to seek information about their treatment (*P*=.02).

Regarding understanding internet health information, patients replied that it seemed easy always (49/472, 10.4%) or sometimes (113/472, 23.9%). For the question as to whether participants trusted the information they found on the internet, overall, 13.6% (64/472) responded no, but it was a statistically more frequent reply among young patients (*P*<.001) and homemakers and retired people had less trust than others (*P*<.001). Patients with secondary or university education reported more frequently than other groups that they trusted the information they found on the internet depending on the website (*P*<.001).

Overall, 16.7% (79/472) of patients turned to the internet for information about their disease or treatment before their doctor’s appointment, an action that was more common in younger patients (*P*=.02) and in students (*P*=.049). After doctor appointment, 18.8% (89/472) of patients always referred to the internet and 31.6% (150/472) only when they had doubts, being more frequent among student patients (*P*=.003) and those with secondary or university education (*P*=.002).

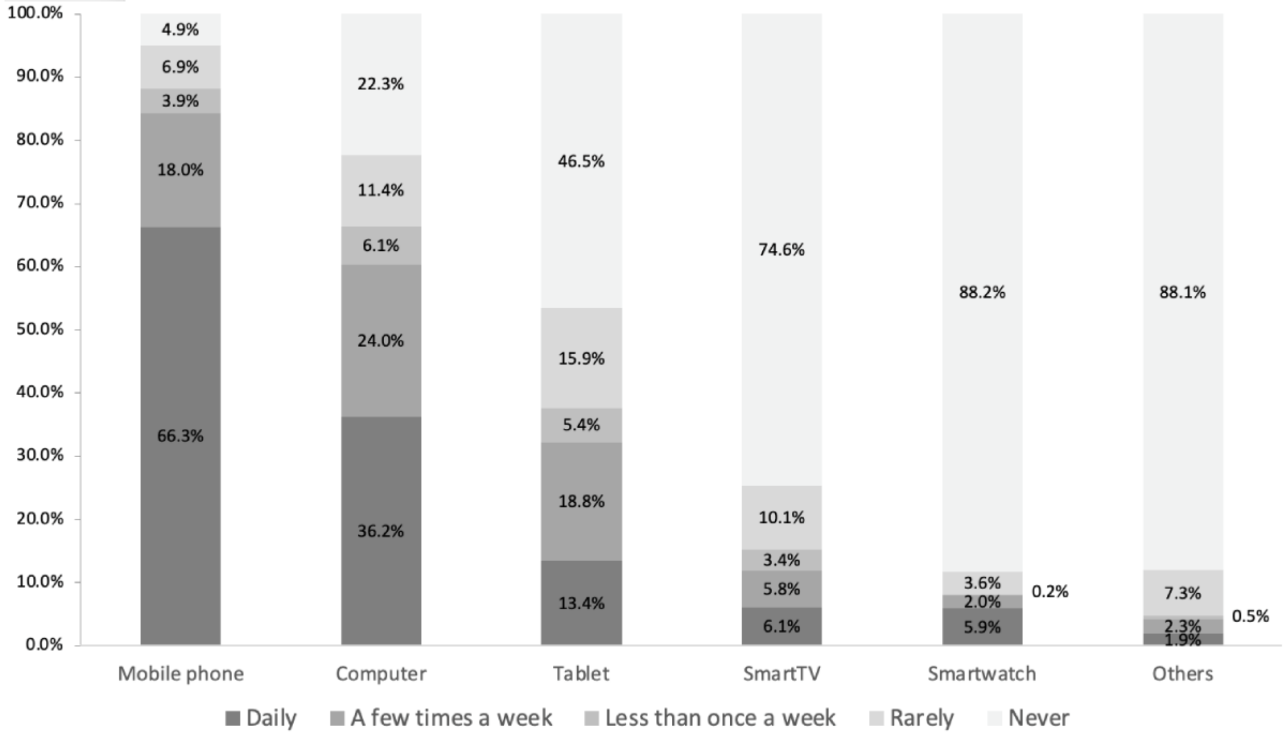
**Table 2.** Information technology use for health-related information (n=472).

| Use of information and communication technologies | Value, n (%) |
|---|---|
| **Which of the following terms have you heard before?[a]** | |
| App | 359 (76.1) |
| Smartphone | 347 (73.5) |
| Information and communication technologies | 178 (37.7) |
| eHealth | 111 (23.5) |
| Mobile health | 89 (18.9) |
| Wearable | 83 (17.6) |
| I have not heard any of these terms before | 53 (11.2) |
| No response | 6 (1.3) |
| **Which devices do you use to look for information on the internet?[a]** | |
| Mobile phone | 433 (91.7) |
| Desktop or laptop computer | 308 (65.3) |
| Tablet | 199 (42.2) |
| Television with internet (SmartTV) | 79 (16.7) |
| Smartwatch | 34 (7.2) |
| Other | 5 (1.1) |
| I do not look for information on the internet | 17 (3.6) |
| No response | 3 (0.6) |
| **Are you interested in staying informed about health-related matters?** | |
| Yes | 410 (86.9) |
| No | 36 (7.6) |
| No response | 26 (5.5) |
| **Where do you search for information on health?[a]** | |
| Internet | 334 (70.8) |
| Health professionals | 318 (67.4) |
| People close to me (friends, relatives, workmates, etc) | 116 (24.6) |
| Newspapers, magazines, pamphlets | 89 (18.9) |
| Apps | 71 (15.0) |
| Other | 12 (2.5) |
| No response | 27 (5.7) |
| **If you use the internet to search for health information, which types of website do you use?[a]** | |
| Google | 347 (73.5) |
| Medical societies | 114 (24.2) |
| Patient associations | 87 (18.4) |
| YouTube | 72 (15.3) |
| I do not search for medical information on the internet | 64 (13.6) |
| Facebook | 49 (10.4) |
| Blogs | 37 (7.8) |
| Twitter | 27 (5.7) |
| Other | 15 (3.2) |
| Other social networks | 10 (2.1) |

| Use of information and communication technologies | Value, n (%) |
|---|---|
| No response | 29 (6.1) |
| **For what purposes do you search for health information?[a]** | |
| Disease prevention, healthy lifestyle, health care | 239 (50.6) |
| To find information about the treatment prescribed by my doctor | 215 (45.6) |
| To find symptoms and learn about potential diseases | 132 (28.0) |
| To find information about medical centers or health professionals | 126 (26.7) |
| To find information about alternative/complementary medicines (herbal products, acupuncture, etc) | 65 (13.8) |
| To get in contact with other people with health problems like mine | 65 (13.8) |
| Other | 21 (4.4) |
| No response | 67 (14.2) |
| **Is it easy to understand the health information you find on the internet?** | |
| Usually | 232 (49.1) |
| Sometimes | 113 (23.9) |
| Always | 49 (10.4) |
| Never | 20 (4.2) |
| No response | 58 (12.3) |
| **Do you trust the health information you find on the internet?** | |
| Depends on the website | 321 (68.0) |
| No | 64 (13.6) |
| Yes | 43 (9.1) |
| No response | 44 (9.3) |
| **Do you look up information on the internet about your disease or treatment BEFORE going to your doctor's appointment?** | |
| No | 367 (77.8) |
| Yes | 79 (16.7) |
| No response | 26 (5.5) |
| **Do you look up information on the internet about your disease or treatment AFTER going to your doctor's appointment?** | |
| No | 210 (44.5) |
| Only if I have still got doubts about something | 150 (31.8) |
| Yes | 89 (18.8) |
| No response | 23 (4.9) |

[a] Multiple choice question.

**Figure 1.** Frequency with which patients used different information and communication technologies to search for information on the internet.



## Patient Expectations About mHealth

Table 3 shows the results for preferences regarding the use of mHealth. Overall, 92.2% (435/472) of patients had a smartphone, and 42.8% (202/472) had installed at least 1 health app. We observed that most (199/472) of the patients had found the health-related apps for their will. Concerning mobile phone use, 73.9% (349/472) of patients used them to access the internet, and 66.5% (314/472) to use apps, among other functions.

Figure 2 shows the reasons why our patients used health apps and their use preferences. We found that patients used health-related apps mainly to manage appointments with health centers, hospitals, or health professionals (222/472, 47.0%) and to obtain information about diseases and treatments (119/472, 25.2%).

We found that younger patients used apps to obtain information about disease prevention, health problems, and improve their lifestyle (*P*=.05) and to record and monitor their symptoms (*P*=.01) with a greater frequency than the other sociodemographic groups. Patients with secondary or university education were statistically more inclined than other patient groups to use apps to record and monitor their symptoms (*P*<.04), record their medication (*P*=.04), and manage their appointments (*P*=.03). Unemployed patients and homemakers used apps more than those in other occupations to get in contact with other patients (*P*=.03). Furthermore, women used apps to obtain emotional support more frequently than men (*P*=.02).

When ask patients about what they would like a health app to help, most of them answered to manage appointments (281/472, 59.5%), to obtain information about their diseases and treatments (274/472, 58.1%), to get in contact with health professionals (250/472, 53.0%), and for remote monitoring by health professionals (226/472, 47.9%). We observed that women than men would like to use apps more to get emotional support (*P*=.001).

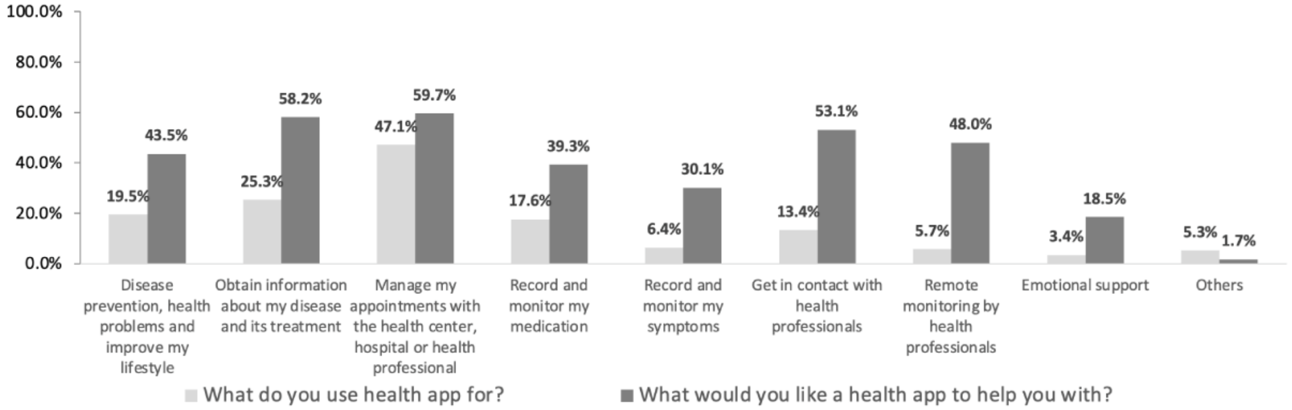
Overall, 90.0% (425/472) of patients reported they would use an app to manage their IMID if their health professionals recommended it, and 58.0% (274/472) patients would pay or probably be willing to pay for it, being more frequent in patients with secondary or university education (*P*<.001).

**Table 3.** Use preferences for health apps (n=472).

| Use preferences for health apps | Value, n (%) |
|---|---|
| **Is your mobile phone a smartphone?** | |
| Yes | 435 (92.2) |
| No | 15 (3.2) |
| No response | 22 (4.7) |
| **Do you have health-related apps installed?** | |
| No | 250 (53.0) |
| Yes | 202 (42.8) |
| No response | 20 (4.2) |
| **How did you find the health-related apps you use?**[a] | |
| Prescription or medical advice | 76 (16.1) |
| On my own | 199 (42.2) |
| On the recommendation of a family member or friend | 73 (15.5) |
| No response | 180 (38.1) |
| **What do you use your mobile phone for?**[a] | |
| Normal phone use (calls, messages, photos/videos, etc) | 448 (94.9) |
| To access the internet | 349 (73.9) |
| Schedule planner and alarms | 323 (68.4) |
| To use apps | 314 (66.5) |
| Social networks | 245 (51.9) |
| No response | 19 (4.0) |
| **How would you like to communicate with your health professional?**[a] | |
| Telephone | 353 (75.0) |
| Videoconference | 202 (42.8) |
| Email | 201 (42.6) |
| Apps | 179 (37.9) |
| Website | 25 (5.3) |
| Blogs | 4 (0.8) |
| Social networks | 16 (3.4) |
| No response | 36 (7.6) |
| **Would you use an app if your health professional recommended it?** | |
| No | 21 (4.4) |
| Yes | 425 (90.0) |
| No response | 26 (5.5) |
| **Would you download a health-related app if you had to pay approximately €2.15 (US $2.15)?** | |
| No | 171 (36.2) |
| Probably | 174 (36.9) |
| Yes | 100 (21.2) |
| No response | 27 (5.7) |

[a] Multiple choice question.

**Figure 2.** Reasons why patients used a health app, and things they wish the app could help them with.



## Discussion

### Principal Findings

We have analyzed the needs and interests of a large cohort of IMID patients in ICTs as a means of managing their disease. We found that the most frequent ICTs used by far was the mobile phone (91.7%) and the most frequent source of information was the internet (70.8%; mainly in Google), ahead of health care professionals.

We showed that patients searched for health information on the internet mainly for disease prevention, healthy lifestyles, health care, and treatments prescribed by their doctors. Most (67.9%) of the patients trusted the information they found on the internet but that depended on the website.

Regarding the use of health apps, 42.8% of patients had health-related apps installed on their mobile phones, with only 16.1% found by medical advice. The most interesting features that patients expected to find in an app were appointment management, advice on disease and treatment management, and personalized communication with health professionals. Most (425/472, 90.0%) of the patients would use an app if their health professional recommended it and more than half (274/472, 58.1%) of patients would be willing to pay for it.

### Comparison With Prior Work

### *Patients Preferences in the Use of Apps*

Digitization of health care and ICT has gained momentum in recent years, mainly sparked by the COVID-19 pandemic [22], changing the practice of medicine and the way in which health information about health and manage diseases is accessed [20,23,24]. However, data published about the use of mHealth technologies in IMID patients is focused on interventions regardless of the patient’s perspective, which is needed for further validation and improvement of these technologies [9-11,13,15]. To the best of our knowledge, this is the first survey to identify the frequency of use and needs and interests in ICTs as a mean of managing their disease in a multidisciplinary cohort of IMID patients, including patients with rheumatology disease, inflammatory bowel diseases, and psoriasis among others, carried out in a comprehensive care center for these patients. The characteristics of the questionnaire, which contains transversal questions, facilitate its extrapolation to other pathologies. Our colleagues had used a very similar questionnaire in hematology-oncology patients to understand the ICT use profile and identify their needs and interest [25].

### *ICT Use for Health-Related Information*

The way in which people access health information is changing. As observed by Knitza et al [18], the internet was the most frequent source of health information, ahead of health care professionals (70.8% vs 67.4%), contrary to what is traditionally described, where more patients have tried to physically obtain medical information from ordinary health professionals [20,25,26]. Regarding the reasons why patients seek health information, most of the patients showed interest in disease prevention (50.5%) and in their treatments (45.6%), although this result was lower compared with other authors (67%-80%) [17,18]. A considerable percentage of patients (47.4%, 35.0%, and 13.8% reported by Magnol et al [17], Knitza et al [18], and in this study, respectively) seek information to get in contact with other people with similar problems. However, when patients were asked about what they would like a health app to do, they showed less interest in getting emotional support or direct exchange such as chats with peers with the same disease [18]. Health professionals can encourage patients to enroll in patient associations for a holistic approach to managing their condition due to its beneficial effects [27,28], and apps could be the platform for it.

The high complexity of managing IMIDs and their treatments, as well as the limited health care resources in many cases, causes patients to have many doubts that are not resolved by health professionals. In our study, more than half (239/472, 50.6%) of the patients consulted the internet after a medical visit. Knitza et al [18] showed that 75% of patients had previously used the internet to obtain health information during the last 3 months prior to the clinical visit. The internet can be an unreliable source of information if they don’t have the ability to make critical use of it [20,26,29]. In our study, only a small percentage of the patients found the information on the internet always easy to understand (10.4%). In addition, 13.4% did not trust the information they found. This highlights the need to establish legislation that regulates this aspect and the role of health professionals to provide guidance on where and how to look for health information [21,30].

### *Patient Expectations About mHealth*

Regarding the use of mHealth, we found that 42.8% of our patients used health-related apps, a higher rate compared with 21.6% showed by Magnol et al [17] or 4.1% reported by Knitza et al [18] in rheumatology patients. This difference could be explained by the younger age of the patients included in our study or the heterogeneity of our patients with different IMIDs. Other chronic diseases has been associated with higher eHealth use [31-33]. In our study, most of the patients (42.2%) found the app on their own. This highlights once again the importance of health professionals in advising on the use of ICTs.

In general, the main interests of patients with apps were information about medications and diseases. Regarding the communication with health care professionals, our patients preferred the more traditional means of communication such as telephone, videoconferencing, and email, as also reported by Knitza et al [18], who observed that patients preferred to receive medical information on paper. However, we showed that one of the app functions that patients would like the most was to get in contact with health professionals (53.0%) and for remote monitoring by health professionals (47.9%). Moreover, 37.9% of patients would use an app as a means of communication with health care professionals. Finally, 90% of our patients would use an app to manage their IMID if their health professionals recommended it, a higher result than showed by Magnol et al [17] (69.9%). The widespread access and use of the mobile phone by patients can help health care professionals increase patient monitoring while improving patient convenience, especially for those who are functionally incapacitated or who live far away [34,35]. This fact, along with the significant demonstrated benefits of apps in terms of clinical care [12,33,36,37], cost-effectiveness [13,38], and versatility of its features (such as including patient-reported outcomes) [39-41], make apps a perfect tool for complement and improve the standard clinical care of patients with IMID.

## Limitations and Strengths

Our study has several limitations related to the characteristics of studies under real clinical practice. First, our study is limited by its single-center design. Consequently, these results might not be obtained in centers or countries with different characteristics. Second, the cross-sectional design, self-reported data, and sampling method can affect extrapolation of the results. However, a strength of this study is its multidisciplinary design with a systematic inclusion of large numbers of consecutive patients with different IMIDs, including rheumatology diseases, inflammatory bowel disease, and psoriasis among others, which resulted in a representative sample of patients with IMID. Third, there is great variability in the price of health-related app that can influence the response of patients to these new technologies depending on the cost indicated in the survey. A review showed that most apps were free, and specifically 82.2% of apps for IMID patients were free [42]. However, in IMID apps requiring payment, the cost is slightly higher (€9.10, or US $9.10) than the average observed in other reviews (€0.90-4.20, or US $0.90-4.20) [42]. Our patients are willing to pay for a health app if their health care professional recommends it.

## Conclusion

IMID patients were very interested in finding health-related information via ICTs, especially using smartphones and apps recommended by health professionals. We could successfully identify unmet needs and patient priorities, with appointment management, advice on disease and treatment management, and personalized communication with health professionals the most app features in which patients were interested. These ICTs, presented as tools that cover these needs, are used as both a source of information and a new communication channel between patients and health professionals. Health professionals should play an essential role in recommending and validating these tools to ensure they are of high quality. Therefore, our results may help in developing possible technological solutions that favor the empowerment of patients and guiding ITCs in routine IMID care.

## Authors' Contributions

ECV, RRJ, VEV, AAL, MELM, AHA, and MS were involved in conceptualizing or designing the study. ECV, RRJ, AAL, and MELM were involved in data collection. ECV, RRJ, and VEV were involved in data analysis and interpretation. ECV, RRJ, and VEV were responsible for drafting the manuscript. ECV, RRJ, VEV, AAL, MELM, CMG, LM, OB, LIF, CLR, AHA, and MS provided critical revisions to the manuscript. All authors have read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 108 KB - jmir_v24i9e37445_app1.docx ]

## References

1. Kuek A, Hazleman BL, Ostör AJK. Immune-mediated inflammatory diseases (IMIDs) and biologic therapy: a medical revolution. Postgrad Med J 2007 Apr;83(978):251-260 [FREE Full text] [doi: 10.1136/pgmj.2006.052688] [Medline: 17403952]

2. Strand V, Singh JA. Newer biological agents in rheumatoid arthritis: impact on health-related quality of life and productivity. Drugs 2010;70(2):121-145. [doi: 10.2165/11531980-000000000-00000] [Medline: 20108988]
3. Coto P, Riestra S, Rozas P, Señaris A, Queiro R. Improving the standard of care for patients with spondyloarthritis-related immune inflammatory conditions: results of a Delphi study and proposal for early detection. Ther Adv Chronic Dis 2020;11:2040622320904295 [FREE Full text] [doi: 10.1177/2040622320904295] [Medline: 32110287]
4. Feagins LA, Iqbal R, Spechler SJ. Case-control study of factors that trigger inflammatory bowel disease flares. World J Gastroenterol 2014 Apr 21;20(15):4329-4334 [FREE Full text] [doi: 10.3748/wjg.v20.i15.4329] [Medline: 24764669]
5. Kool MB, van Middendorp H, Lumley MA, Schenk Y, Jacobs JWG, Bijlsma JWJ, et al. Lack of understanding in fibromyalgia and rheumatoid arthritis: the Illness Invalidation Inventory (3*I). Ann Rheum Dis 2010 Nov 24;69(11):1990-1995. [doi: 10.1136/ard.2009.123224] [Medline: 20498203]
6. Armstrong AW, Read C. Pathophysiology, clinical presentation, and treatment of psoriasis: a review. JAMA 2020 May 19;323(19):1945-1960. [doi: 10.1001/jama.2020.4006] [Medline: 32427307]
7. Grady A, Yoong S, Sutherland R, Lee H, Nathan N, Wolfenden L. Improving the public health impact of eHealth and mHealth interventions. Aust N Z J Public Health 2018 Apr;42(2):118-119. [doi: 10.1111/1753-6405.12771] [Medline: 29384248]
8. García CF, de San Pedro M, Ledesma CA. Las TICs y la gestión de pacientes crónicos. Sociedad Española Informática Salud 2014;105:11-13 [FREE Full text] [doi: 10.1353/nef.2011.0046]
9. Kataria S, Ravindran V. Digital health: a new dimension in rheumatology patient care. Rheumatol Int 2018 Nov;38(11):1949-1957. [doi: 10.1007/s00296-018-4037-x] [Medline: 29713795]
10. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. Lancet 2017 Sep 02;390(10098):959-968. [doi: 10.1016/S0140-6736(17)31327-2] [Medline: 28716313]
11. Dahy A, El-Qushayri AE, Mahmoud AR, Al-Kelany TA, Salman S. Telemedicine approach for psoriasis management, time for application? A systematic review of published studies. Dermatol Ther 2020 Nov;33(6):e13908. [doi: 10.1111/dth.13908] [Medline: 32592525]
12. Yin AL, Hachuel D, Pollak JP, Scherl EJ, Estrin D. Digital health apps in the clinical care of inflammatory bowel disease: scoping review. J Med Internet Res 2019 Aug 19;21(8):e14630 [FREE Full text] [doi: 10.2196/14630] [Medline: 31429410]
13. Jackson BD, Gray K, Knowles SR, De Cruz P. eHealth technologies in inflammatory bowel disease: a systematic review. J Crohns Colitis 2016 Sep;10(9):1103-1121. [doi: 10.1093/ecco-jcc/jjw059] [Medline: 26928960]
14. Gossec L, Baraliakos X, McInnes I, Kerschbaumer A, de Wit M, Dougados M, et al. Response to: 'Comment on: 'EULAR recommendations for the management of psoriatic arthritis with pharmacological therapies: 2019 update' by Gossec ' by Wei. Ann Rheum Dis 2022 Aug 01;81(8):e139. [doi: 10.1136/annrheumdis-2020-218456] [Medline: 32873555]
15. van Riel P, Alten R, Combe B, Abdulganieva D, Bousquet P, Courtenay M, et al. Improving inflammatory arthritis management through tighter monitoring of patients and the use of innovative electronic tools. RMD Open 2016;2(2):e000302 [FREE Full text] [doi: 10.1136/rmdopen-2016-000302] [Medline: 27933206]
16. Romero-Jimenez R, Escudero-Vilaplana V, Chamorro-De-Vega E, Ais-Larisgoitia A, Lobato Matilla ME, Herranz-Alonso A, et al. The characteristics and functionalities of mobile apps aimed at patients diagnosed with immune-mediated inflammatory diseases: systematic app search. J Med Internet Res 2022 Mar 04;24(3):e31016 [FREE Full text] [doi: 10.2196/31016] [Medline: 35254286]
17. Magnol M, Eleonore B, Claire R, Castagne B, Pugibet M, Lukas C, et al. Use of eHealth by patients with rheumatoid arthritis: observational, cross-sectional, multicenter study. J Med Internet Res 2021 Jan 29;23(1):e19998 [FREE Full text] [doi: 10.2196/19998] [Medline: 33512320]
18. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]
19. Con D, Jackson B, Gray K, De Cruz P. eHealth for inflammatory bowel disease self-management—the patient perspective. Scand J Gastroenterol 2017 Sep;52(9):973-980. [doi: 10.1080/00365521.2017.1333625] [Medline: 28598210]
20. Opiniones y expectativas de los ciudadanos sobre el uso y aplicación de las TIC en el ámbito sanitario. Los ciudadanos ante la e-Sanidad. Ontsi.red. URL: https://www.ontsi.es/sites/ontsi/files/informe_ciudadanos_esanidad.pdf [accessed 2022-08-24]
21. Collado-Borrell R, Escudero-Vilaplana V, Ribed-Sánchez A, Ibáñez-García S, Herranz-Alonso A, Sanjurjo-Sáez M. Smartphone applications for cancer patients: what we know about them? Farm Hosp 2016 Jan 01;40(1):25-35 [FREE Full text] [doi: 10.7399/fh.2016.40.1.8993] [Medline: 26882831]
22. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020 Dec 01;22(12):e24087. [doi: 10.2196/24087] [Medline: 33147166]
23. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]

24. Aitken M, Lyle J. Patient adoption of mHealth: use, evidence and remaining barriers to mainstream acceptance. URL: https://fsc-ccf.ca/references/patient-adoption-of-mhealth-use-evidence-and-remaining-barriers-to-mainstream-acceptance/ [accessed 2022-08-24]
25. Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, et al. Oncology patient interest in the use of new technologies to manage their disease: cross-sectional survey. J Med Internet Res 2018 Oct 23;20(10):e11006 [FREE Full text] [doi: 10.2196/11006] [Medline: 30355554]
26. Mira JJ, Pérez-Jover V, Lorenzo S. [Navigating on the internet in search of health information: all that glitters is not gold]. Aten Primaria 2004 Apr 30;33(7):391-399 [FREE Full text] [doi: 10.1016/s0212-6567(04)78892-7] [Medline: 15117635]
27. Garrido-Cumbrera M, Chacón-García J, Navarro-Compán V, Gratacós J, Sanz-Gómez S, Collantes-Estevez E, Atlas working group. Does belonging to a patient association is of help for patients with axial spondyloarthritis? Results from the Atlas Survey. Curr Rheumatol Rep 2020 May 14;22(6):22. [doi: 10.1007/s11926-020-00897-5] [Medline: 32410101]
28. González-de Paz L, Ribera M, Gratacós-Masmitjà J, Ros S, Blanch Mur C, Alfonso-Zamora S. The Spanish Psoriasis Patients' Association according to its members: cross-sectional descriptive study of members' opinions of the association's functions. Actas Dermosifiliogr 2016 Nov;107(9):751-757 [FREE Full text] [doi: 10.1016/j.ad.2016.06.003] [Medline: 27422558]
29. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
30. Krebs P, Duncan DT. Health app use among us mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
31. Eysenbach G. The impact of the internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [doi: 10.3322/canjclin.53.6.356] [Medline: 15224975]
32. Hansen AH, Broz J, Claudi T, Årsand E. Relations between the use of electronic health and the use of general practitioner and somatic specialist visits in patients with type 1 diabetes: cross-sectional study. J Med Internet Res 2018 Nov 07;20(11):e11322 [FREE Full text] [doi: 10.2196/11322] [Medline: 30404766]
33. Zhang Y, Lauche R, Sibbritt D, Olaniran B, Cook R, Adams J. Comparison of health information technology use between American adults with and without chronic health conditions: findings From The National Health Interview Survey 2012. J Med Internet Res 2017 Oct 05;19(10):e335 [FREE Full text] [doi: 10.2196/jmir.6989] [Medline: 28982644]
34. Sargious A, Lee SJ. Remote collection of questionnaires. Clin Exp Rheumatol 2014;32(5 Suppl 85):S168-S172. [Medline: 25365110]
35. The Mobile Economy: 2019. London: GSM Association; 2019. URL: https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/07/GSMA_MobileEconomy2021_3.pdf [accessed 2022-01-15]
36. Kruse CS, Beane A. Health information technology continues to show positive effect on medical outcomes: systematic review. J Med Internet Res 2018 Feb 05;20(2):e41 [FREE Full text] [doi: 10.2196/jmir.8793] [Medline: 29402759]
37. Domogalla L, Beck A, Schulze-Hagen T, Herr R, Benecke J, Schmieder A. Impact of an eHealth smartphone app on the mental health of patients with psoriasis: prospective randomized controlled intervention study. JMIR Mhealth Uhealth 2021 Oct 25;9(10):e28149 [FREE Full text] [doi: 10.2196/28149] [Medline: 34431478]
38. de Jong MJ, Boonen A, van der Meulen-de Jong AE, Romberg-Camps MJ, van Bodegraven AA, Mahmmod N, et al. Cost-effectiveness of telemedicine-directed specialized vs standard care for patients with inflammatory bowel diseases in a randomized trial. Clin Gastroenterol Hepatol 2020 Jul;18(8):1744-1752. [doi: 10.1016/j.cgh.2020.04.038] [Medline: 32335133]
39. Wicks P, Stamford J, Grootenhuis MA, Haverman L, Ahmed S. Innovations in e-health. Qual Life Res 2014 Feb;23(1):195-203 [FREE Full text] [doi: 10.1007/s11136-013-0458-x] [Medline: 23852096]
40. Williet N, Sandborn WJ, Peyrin-Biroulet L. Patient-reported outcomes as primary end points in clinical trials of inflammatory bowel disease. Clin Gastroenterol Hepatol 2014 Aug;12(8):1246-1256. [doi: 10.1016/j.cgh.2014.02.016] [Medline: 24534550]
41. Peyrin-Biroulet L, Sandborn W, Sands BE, Reinisch W, Bemelman W, Bryant RV, et al. Selecting therapeutic targets in inflammatory bowel disease (STRIDE): determining therapeutic goals for treat-to-target. Am J Gastroenterol 2015 Sep;110(9):1324-1338. [doi: 10.1038/ajg.2015.233] [Medline: 26303131]
42. Romero-Jimenez R, Escudero-Vilaplana V, Chamorro-De-Vega E, Ais-Larisgoitia A, Lobato Matilla ME, Herranz-Alonso A, et al. The characteristics and functionalities of mobile apps aimed at patients diagnosed with immune-mediated inflammatory diseases: systematic app search. J Med Internet Res 2022 Mar 04;24(3):e31016 [FREE Full text] [doi: 10.2196/31016] [Medline: 35254286]

## Abbreviations

**ICT:** information technology  
**IMID:** immune-mediated inflammatory disease  
**mHealth:** mobile health

*Edited by G Eysenbach; submitted 21.02.22; peer-reviewed by Á Narrillos Moraza, S García-Sánchez; comments to author 28.03.22; revised version received 29.04.22; accepted 18.07.22; published 13.09.22.*

*Please cite as:*
*Chamorro-de-Vega E, Romero-Jiménez R, Escudero-Vilaplana V, Ais-Larisgoitia A, Lobato Matilla ME, González CM, Menchén L, Baniandrés O, Ibares-Frias L, Lobo-Rodríguez C, Herranz-Alonso A, Sanjurjo M*
*Information and Communication Technologies in Patients With Immune-Mediated Inflammatory Diseases: Cross-sectional Survey*
*J Med Internet Res 2022;24(9):e37445*
*URL: https://www.jmir.org/2022/9/e37445*
*doi:10.2196/37445*
*PMID:36099018*

Original Paper

# Twitter Usage Among Physicians From 2016 to 2020: Algorithm Development and Longitudinal Analysis Study

Keisuke Nakagawa[1,2], MD; Nuen Tsang Yang[3], MS; Machelle Wilson[3], PhD; Peter Yellowlees[1], MBBS, MD

[1]Department of Psychiatry and Behavioral Sciences, University of California, Davis School of Medicine, Sacramento, CA, United States
[2]Digital CoLab, Innovation Technology, University of California, Davis Health, Sacramento, CA, United States
[3]Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, Sacramento, CA, United States

**Corresponding Author:**
Keisuke Nakagawa, MD
Department of Psychiatry and Behavioral Sciences
University of California, Davis School of Medicine
2230 Stockton Blvd
Sacramento, CA, 95817
United States
Phone: 1 916 397 8941
Email: drknakagawa@ucdavis.edu

## Abstract

**Background:** Physicians are increasingly using Twitter as a channel for communicating with colleagues and the public. Identifying physicians on Twitter is difficult due to the varied and imprecise ways that people self-identify themselves on the social media platform. This is the first study to describe a reliable, repeatable methodology for identifying physicians on Twitter. By using this approach, we characterized the longitudinal activity of US physicians on Twitter.

**Objective:** We aimed to develop a reliable and repeatable methodology for identifying US physicians on Twitter and to characterize their activity on Twitter over 5 years by activity, tweeted topic, and account type.

**Methods:** In this study, 5 years of Twitter data (2016-2020) were mined for physician accounts. US physicians on Twitter were identified by using a custom-built algorithm to screen for physician identifiers in the Twitter handles, user profiles, and tweeted content. The number of tweets by physician accounts from the 5-year period were counted and analyzed. The top 100 hashtags were identified, categorized into topics, and analyzed.

**Results:** Approximately 1 trillion tweets were mined to identify 6,399,146 (<0.001%) tweets originating from 39,084 US physician accounts. Over the 5-year period, the number of US physicians tweeting more than doubled (ie, increased by 112%). Across all 5 years, the most popular themes were general health, medical education, and mental health, and in specific years, the number of tweets related to elections (2016 and 2020), Black Lives Matter (2020), and COVID-19 (2020) increased.

**Conclusions:** Twitter has become an increasingly popular social media platform for US physicians over the past 5 years, and their use of Twitter has evolved to cover a broad range of topics, including science, politics, social activism, and COVID-19. We have developed an accurate, repeatable methodology for identifying US physicians on Twitter and have characterized their activity.

***(J Med Internet Res 2022;24(9):e37752)*** doi:10.2196/37752

**KEYWORDS**

social media; internet; health informatics; internet use; public health; Twitter; physician

## Introduction

Twitter has become a popular communication tool, with almost 200 million daily active users and more than 500 million tweets being sent out every day [1]. Around the world, people have used Twitter as a platform for sharing information, expressing opinions, and engaging in social movements. It has democratized communication, allowing everyday citizens to have a voice at the same potential scale as the voices of global leaders, politicians, and celebrities. The speed at which information is shared on Twitter has transformed the way information spreads and how communities create social movements. Studies have shown that Twitter is one of the best sources of real-time information, outpacing traditional media outlets for reporting natural disasters, crimes, and major events [2-6].

Over the past 15 years, physicians have started to use Twitter as a platform for sharing scientific information, frontline experiences, and opinions on various topics [7-12]. Studies have shown that scientific papers obtain more visibility when shared on Twitter, and impact metrics have started to incorporate social media data as part of their calculations [13]. During the COVID-19 pandemic, Twitter has given unprecedented visibility to clinicians working on the front lines. Clinicians have engaged in public discourse by sharing data and scientific information on Twitter [14-17]. They have also engaged in dialogue with each other, often sharing encouragement, empathy, and compassion during challenging times [18,19].

Studies have already demonstrated the use of Twitter in building early warning systems for adverse drug reactions, understanding discussions about various diseases, and characterizing public perceptions regarding COVID-19 [20-24]. One of the earliest uses of Twitter among physicians was tweeting during conferences, and a number of studies have analyzed what factors make tweets more successful or more likely to be disseminated during medical conferences [9,25]. Recent studies have demonstrated that physician activity on Twitter is a potential predictor of COVID-19 surges [26]. As more physicians join Twitter and engage in dialogue across a wide range of topics, health systems and academic medical centers that train and often employ these physicians need to be aware of these trends to teach them about professionally appropriate use and potential repercussions. Physicians should be aware of the positive and negative consequences of engaging in public discourse on platforms such as Twitter, where content is instantly made public and is often permanent and irreversible [23,27].

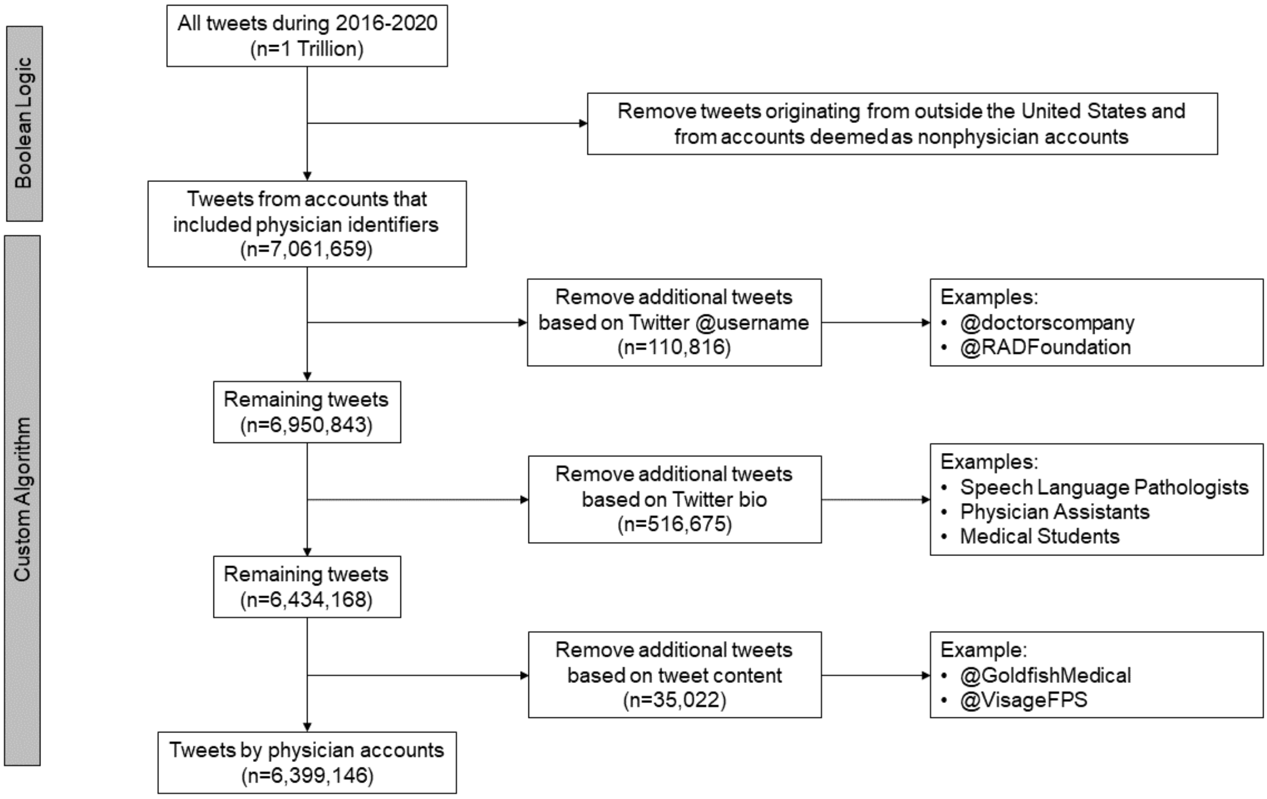
Although some studies have investigated the use of Twitter among physicians in narrow use cases, such as a specific specialty, no study has characterized the overall activity of US physicians on Twitter across multiple years [11]. Part of this can be attributed to the difficulty of identifying physicians on Twitter. Physicians self-report their professional status in varied, inconsistent, and sometimes vague ways, and there are limited methods for verifying and validating the identity of a user based on their profile and tweet content. Furthermore, the challenges of combing through billions of tweets and bios to identify physician accounts can be computationally and logistically intensive. It is for these reasons that a process for identifying US physicians and characterizing their activity has not been published, to our knowledge. Given the widespread use of Twitter by the general public and among physicians, it is becoming increasingly important to understand the use of social media among physicians, particularly with the backdrop of the COVID-19 pandemic. The aims of this study are to describe a reliable and repeatable methodology for identifying US physicians on Twitter and to characterize their use of Twitter and longitudinal trends across multiple years.

## Methods

### Process for Identifying US Physicians on Twitter

We conducted an in silico analysis of physicians' Twitter usage over a 5-year period. We identified physicians by using a 3-stage process (Figure 1) comprising a geographic filter for excluding accounts from outside of the United States, Boolean logic for further identifying US physician accounts based on their usernames and bios, and a custom algorithm for filtering out false-positive accounts. The details of the algorithm developed and the iterative refinement process are described in the following sections.

**Figure 1.** Identification and selection of physician tweets.



## Algorithm Development and Refinement for Physician Identification

Initially, a Boolean logic was developed (Multimedia Appendix 1) to scan all Twitter usernames (ie, *@username*), names (eg, *Jane Doe, M.D.*), and bios for every tweet that was posted during the 5-year study period (2016-2020). The Boolean logic identified all tweets that included some indication of a Twitter user's status as a physician (eg, mentioning *physician* in the bio or including a medical degree, such as *M.D.* or *D.O.*, in the name or username). The Boolean logic included a corpus that contained every specialty and subspecialty listed by the Accreditation Council for Graduate Medical Education and the American Osteopathic Association. A third-party software (Meltwater Explore; Meltwater, Inc) was used to review the user accounts and tweets generated from the Boolean algorithm to ensure the accuracy of the Boolean logic expression. With each iteration, the first 1000 tweets were manually reviewed for any tweets that did not originate from a physician-owned account. Appropriate logic was added to filter out false positives. Examples of false positives are provided in Table 1. This manual review process was repeated until the false-positive rate decreased to less than 5% in a 1000-tweet sample.

**Table 1.** Examples of false positives and the associated Boolean logic.

| Description | Examples | Boolean logic |
|---|---|---|
| *MD* is used in the username, in the profile, or in both to indicate something other than a medical degree. | • *MD* in "@MDZeki" stands for "Muhammad"<br>• *MD* in "Student living in MD" and "@MDJobs" stands for "Maryland" | • Only identify tweets that are in English and from the United States<br>• Only include usernames that have ",M.D.," ",MD," or ", MD"<br>• Remove all accounts that include *jobs* in the username or profile |
| *DO* is used in the username, in the profile, or in both to indicate something other than a medical degree. | • *DO* in "@JUSTDOIT" does not indicate a DO medical degree | • Only include names that have ",D.O." |
| *Doctor* is used broadly without specifying an MD or DO degree. | • @DoctorJones<br>• @aprokodoctor<br>• @DoctorsEMres | • The Boolean logic did not select handles with *Doctor*; the user would have had to indicate their status as a physician through more specific ways |
| Premedical and medical students who include *physician* or *doctor* in their username or profile. | • Profile includes "Aspiring Infectious Disease Physician"<br>• Profile includes "aspiring physician working in epi & vaccine research"<br>• Profile includes "future physician" or "Future Physician Assistant" | • Remove all accounts that include *aspiring* or *future* in the profile |

## Data Acquisition

Once the Boolean logic was finalized, all tweets that were written in English; originated from the United States; were posted between January 1, 2016, to December 31, 2020; and met the Boolean criteria were acquired from a third-party vendor (Meltwater, Inc, San Francisco, California).

## Algorithm Refinement

The data were imported into RStudio (Version 1.2.5033; RStudio, PBC), where additional rules were programmed in R (R Foundation for Statistical Computing) and applied to filter out nonphysician Twitter accounts. The next phase of algorithm refinement involved filtering out more nuanced cases that the Boolean logic could not catch. Accounts that used *MD* to indicate the state of Maryland; accounts of physician recruiting services that used variations of *MD* and *jobs* in their usernames and profiles; and premedical and medical students who used *aspiring MD*, *aspiring physician*, *future physician*, or similar terms in their profile were filtered out by using logic coded in R.

The algorithm was further refined by identifying patterns in the tweet text itself and then applying filtering logic to the accounts of those tweets. For example, to filter out more physician recruitment accounts, tweets that used the hashtags *#job* or *#jobs* in the text were identified. Among these tweets, any accounts that included the word *medical* or *surgery* in the account name were eliminated (eg, *Goldfish Medical* or *Visage Facial Plastic Surgery*). This 2-step process of screening accounts—analyzing the tweet text first and then identifying false-positive usernames and names—was used to filter out the final batch of nonphysician accounts.

## Algorithm Accuracy and Validation

We used Twitter's verified accounts (ie, those with a blue badge) as the validation set to test the accuracy of the algorithm. Twitter uses a rigorous verification process to confirm the identities of verified account users, which includes checking for an official government-issued identification document or an official email address [28].

All of the tweets that originated from a verified Twitter account were separated from the larger 5-year data set. A team manually reviewed each unverified account by visiting the associated Twitter profiles and searching the internet to confirm the details of a user's status as a physician (eg, checking the website of their stated institution for their profile). A false-positive error rate was calculated based on the number of verified accounts that belonged to Twitter users who were not actually physicians. In order to calculate the false-positive rate of unverified accounts, around 5% (1978/43,328, 4.6%) of the unverified accounts in the data set were randomly selected. The accounts were manually reviewed by the team to calculate a false-positive rate for the unverified physician accounts identified by the algorithm.

## Hashtag and Topic Analysis

The top 100 hashtags used by US physicians were identified, tallied, and ranked for each year. Each tweet was tokenized, stemmed, and normalized, and stop words were removed. This was done by using the following R packages: *stringi*, *stringr*, and *tokenizers*. All words that followed a pound sign were identified and separately stored within a data frame in RStudio. For every tweet, each hashtag was counted and tallied across all tweets for each year (2016-2020). Researchers (KN and NTY) reviewed the list and grouped any terms that were related. For example, the COVID-19 topic was comprised of hashtags such as *#covid19*, *#coronavirus*, *#sarscov2*, *#covid19coronavirus*, and other related hashtags. Further, both the *#blacklivesmatter* and *#blm* hashtags were used for the Black Lives Matter topic's tally and rank.

We analyzed hashtags in two ways. First, we created word clouds for a visual representation of Twitter activity. A word cloud was generated for each year of the study period; words were color-coded based on their frequency of use. Word clouds were generated by using the following R packages: *wordcloud*, *wordcloud2*, and *ggplot2*. Second, we identified the top 100 hashtags and categorized them into broad thematic areas. In order to analyze thematic trends between years, hashtags were organized into the following themes: general health care, COVID-19 and public health, politics, social activism, mental health and well-being, health technology, conferences, patient groups, specialties and subspecialties, medical conditions and procedures, and medical education. For the general health care theme, tweets that included generic health care–oriented hashtags like *#healthcare*, *#doctor*, *#medtwitter*, and *#medicine* were combined and tallied, and the proportion of such tweets was calculated as a percentage of all tweets for each year. For the COVID-19 and public health theme, hashtags like *#covid19*, as well as hashtags that conveyed public health messages like *#maskup*, *#socialdistancing*, and *#flattenthecurve*, were included. For the politics theme, any hashtag related to politics and elections, such as *#trump*, *#vote*, *#election2020*, *#tcot* (ie, top conservatives on Twitter), and *#newsmax*, were used [29,30]. Only the top 100 hashtags for each year were considered.

### Ethical Considerations

No application for an ethics review board assessment was submitted, since this study involved third-party data sets with no experimental activities. Therefore, this study was deemed a quality assurance/quality improvement activity [31].

## Results

### Algorithm Accuracy

Of the 216 Twitter-verified accounts that the Boolean algorithm identified as physician accounts, 12 were not physician accounts, resulting in an accuracy rate of 94.4% (204/216). Of the 43,328 unverified accounts in the data set (ie, regular Twitter accounts), 1978 accounts were randomly selected for manual review as part of the 5% validation sample. Of the 1978 randomly-selected unverified accounts, 204 were not physician accounts, resulting in an accuracy rate of 89.7% (1774/1978; 95% CI 88.3%-91%).

### Twitter Activity

The number of US physicians who were active on Twitter increased from 2016 to 2020. The number of unique physician accounts tweeting increased from 12,675 in 2016 to 26,897 in 2020—a 112.2% increase in the number of active physician users over the 5-year period (Table 2). The number of new physician accounts per year also increased from 1711 new accounts in 2016 to 2215 new accounts in 2020—a 19.7% year-over-year increase in the number of new physician accounts created. Although the total number of US physician accounts increased, the total number of tweets varied across the 5-year period, with 1,461,753 total physician tweets in 2016 and 1,338,150 total physician tweets in 2020. The average number of tweets per account decreased from 115.3 tweets per physician account in 2016 to 49.8 tweets per physician account in 2020. These trends were also reflected in a subgroup analysis of verified Twitter accounts.

**Table 2.** Twitter activity among US physicians from 2016 to 2020.

| Twitter activity | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|
| **US physicians on Twitter** | | | | | |
| Tweets per year, n | 1,461,753 | 1,344,911 | 1,205,053 | 1,049,279 | 1,338,150 |
| Unique accounts tweeting, n | 12,675 | 15,633 | 17,934 | 20,584 | 26,897 |
| New accounts tweeting, n | 1711 | 2170 | 1917 | 1760 | 2215 |
| Number of tweets per account, mean | 115.3 | 86.0 | 67.2 | 51.0 | 49.8 |
| **Verified Twitter accounts**[a] | | | | | |
| Tweets per year, n | 57,255 | 56,082 | 49,743 | 51,776 | 77,609 |
| Unique accounts tweeting, n | 107 | 132 | 155 | 171 | 208 |
| New accounts tweeting, n | 2 | 6 | 3 | 8 | 2 |
| Number of tweets per account, mean | 535.1 | 424.9 | 320.9 | 302.8 | 373.1 |

[a] Verified accounts have a blue verified badge next to the Twitter user's name to let people know that an account is authentic and has undergone rigorous verification by Twitter. More information is available on the Twitter Help Center [28].

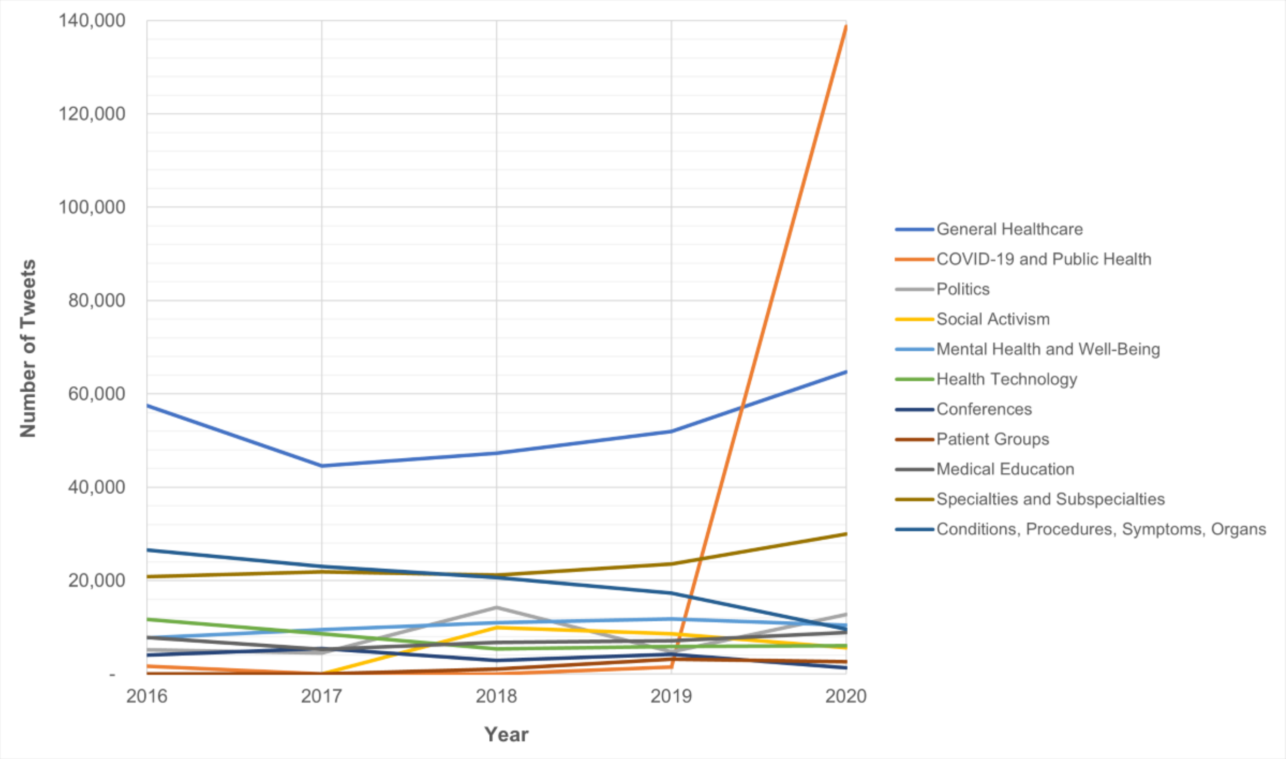
### Hashtag and Topic Analysis

Among US physicians using hashtags in their tweets, *#health*, *#healthcare*, *#doctor*, *#medicine*, *#meded*, *#mentalhealth*, and *#wellness* consistently ranked among the most frequently used hashtags across all 5 years of the study period (Multimedia Appendix 2). In 2016, *#digitalhealth* and *#realdonaldtrump* were trending hashtags, while *#trump*, *#backfiretrump*, *#diversity*, and *#inclusion* were trending in 2018 and 2019. Hashtags related to *#blacklivesmatter* and *#covid19* represented 0.3% (3750/1,291,567) and 10.6% (137,107/1,291,567) of all hashtags used in 2020, respectively. Donald Trump–related hashtags were consistently ranked among the top 50 hashtags from 2016 to 2019, and in 2020, US physicians used COVID-19–related hashtags in 137,107 tweets, making them the most used hashtags during the entire 5-year study period (Multimedia Appendix 2), with *#health* being the second most used hashtag across all 5 years (used in 64,092 tweets).

The word clouds provide a visual representation of the frequency and distribution of the hashtags used each year. The frequency of words was color-coded; gray represents the most frequently used hashtags, green represents the least used hashtags, and other colors (light green, pink, purple, and orange) represent the different levels of frequency in between the two extremes. Multimedia Appendix 3 shows the word clouds for tweets from 2016 to 2019. The word clouds for 2016 to 2019 show a broad distribution of words, as demonstrated by the wide variety of colors and the varying sizes of words in the word clouds (Multimedia Appendix 3). The 2020 word cloud shifts drastically, showing just 1 hashtag in gray (*#COVID-19*) and 1 hashtag in orange (*#medtwitter*); all other hashtags are in green (ie, the least used hashtags).

Once hashtags were organized into themes, the results showed that the general health care topic was the most consistently discussed topic across all 5 years and represented 5% (64,701/1,291,567;2020) of the total tweets in 2020 (Figure 2). In approximately 2% (29,964/1,291,567;2020) of tweets, US physicians discussed specialties and subspecialties, and 0.5% (8830; 1,291,567;2020) of tweets discussed medical education. Public health topics were discussed in less than 0.5% (1478/1,186,835; 2019) of the tweets from 2016 to 2019, but in 2020, US physicians discussed COVID-19 and related public health messages—the most dominant theme across all 5 years of the study period—in over 10% of all tweets.

**Figure 2.** Trending themes by year.



## Discussion

### Study Overview

This is the first longitudinal report of physician activity on Twitter. Physician activity on Twitter has been increasing over the past 5 years, with more physicians creating accounts and joining the platform every year. The diverse range of topics that are discussed on the platform, from social activism and politics to memes, demonstrate that Twitter is becoming a mainstream communication tool for everyday physicians.

The trends in the last 5 years likely indicate that the early adopters phase of Twitter use among physicians is sunsetting, and Twitter is becoming a mainstream platform for physicians [32]. This is supported by the decreasing average number of tweets per user, which shows that the fraction of power users, that is, users with a typically high number of tweets per year, is decreasing in the population of US physician users while use by mainstream physicians, who tend to have a lower number of tweets per year, is increasing. The best example of this occurred in 2020, which was when the content topics drastically shifted toward the topic of COVID-19—a mainstream issue of interest among physicians and the public. Similar trends were seen during the election years (2016 and 2020) and around the time of the Black Lives Matter movement in 2020. If Twitter use among physicians was still in the early adopters phase, we would expect to see less volatile shifts in topics from year to year, since the interests of a niche group would be more stable and consistent.

We also want to acknowledge the role of misinformation and disinformation on the internet. Although this was not the focus of this study, this topic is inextricably linked to the roles and responsibilities that physicians have on Twitter, social media, and the internet. The study period included the 2016 US presidential election and the COVID-19 pandemic, of which both involved the sharing of misinformation and disinformation related to these events [33-36]. Physicians who are active on social media have a disproportionately large influence on the public discourse around science, social movements, and politics, which is supported by the findings in this study. We anticipate that this trend will only continue to increase, placing more importance on studying and understanding the role that physicians play on social media in shaping society's perceptions and opinions.

### Principal Findings

Physicians discuss a broad range of topics on Twitter and are not shy about opining on topics that are relevant to society beyond strictly medical topics. Physicians are unlikely to find an equivalent platform (traditional or social media) where ordinary clinicians' voices can be shared and heard globally in such a public, transparent way. This was illustrated by the outsized influence that physicians like Bob Wachter (@Bob_Wachter), physician influencers like Kevin Pho (@kevinmd), and ordinary clinicians have had by using Twitter during the COVID-19 pandemic. Among the hundreds of millions of tweets shared in 2020 [1], the 137,107 (<0.001%) physicians tweeting about COVID-19 may have had a disproportionately outsized impact on public dialogues and disease perceptions.

US physicians are increasingly using Twitter to teach and share medical education materials. Hashtags like *#meded* and *#foamed* (ie, free open-access medicine) were consistently ranked in the top 10 and top 100 hashtags, respectively (Multimedia Appendix 2). This demonstrates a growing shift toward the democratization of medical education through the sharing of materials, study sheets, and illustrations. This shift toward democratizing medical education is likely to continue, with Twitter and other social media platforms, such as Instagram, playing an increasing role in the future of medical education.

Tweeting during medical conferences continues to be a mainstay of use cases among US physicians, but it has become less prominent relative to more mainstream topics. The American Academy of Pediatrics, American Society of Clinical Oncology, and American College of Cardiology conference hashtags (eg, *#aap18*, *#acc19*, and *#asco20*, respectively) were ranked in the top 100 hashtags across the majority of the 5-year study period (Multimedia Appendix 2). Social media continues to be a powerful tool for disseminating scientific information through medical conferences and scientific journals.

Another major trend in Twitter activity was its growing use for marketing purposes. Hashtags like *#healthcaremarketing*, *#medicalmarketing*, and *#plasticsurgery* were consistently ranked in the top 50 hashtags (Multimedia Appendix 2). During the filtering process, a large group of accounts had to be removed that were associated with physician recruitment services using hashtags like *#jobs* in the profile, in the tweet text, or in both. We also observed individual physicians in certain specialties, such as plastic surgery, cosmetic surgery, and dermatology, using Twitter to market their practice, further reinforcing the fact that Twitter is now accepted as a mainstream platform that physicians can use to communicate directly with patient populations.

Finally, mental health and well-being were consistently ranked in the top 50 topics that US physicians discussed on Twitter. Across a multitude of hashtags, including *#mentalhealth*, *#wellness*, *#depression*, and *#mindfulness*, US physicians used Twitter to engage in dialogue about wellness. These trends may be particularly useful for monitoring physicians' well-being at a local, regional, or national level, since the mental health of clinicians is notoriously challenging to measure on a regular basis. We hypothesize that during stressful events, such as the COVID-19 pandemic, Twitter may offer unprecedented insight into the overall mental health status of physicians, and we are following up this study with one involving a sentiment analysis of all physician tweets that have been posted during the pandemic.

### Challenges, Strengths, and Limitations

During the planning of this study, it was clear to us how challenging it would be to identify physicians on Twitter, hence the lack of publications on this topic in the literature. Twitter was designed to be fast and easy with regard to sign-up and use, and the consequence of this focus on user experience is that there are very few denominators for identifying subgroups within Twitter. The most challenging task of this study was identifying US physician tweets (n=6,399,146) from the trillions of tweets generated worldwide.

Although numerous studies have analyzed the Twitter activity of the entire Twitter population, the complexity of this task increases significantly once subgroups need to be identified and analyzed. We believe that the methodology outlined in this study offers a robust, scalable, and repeatable method of identifying physician accounts on Twitter. We also recognize that this approach is not perfect and hope that this study serves as a foundation for future work to further improve the identification of physician accounts for future analyses.

The computational requirements for mining approximately 1 trillion tweets and analyzing millions of tweets should not be underestimated. Once the data set reaches beyond a few months of Twitter data, the data set is too large for manual assessment. As such, algorithmic approaches to filtering and analyzing the data are recommended. Analyzing this amount of data often exceeds the computational limits of a typical laptop or desktop, and more high-performance graphics processing units or high-performance computing clusters in the cloud are recommended to analyze the data in a reasonable amount of time, particularly if a real-time analysis is desired.

This study had 3 main limitations. First, all physicians who were identified as active users on Twitter were self-reported physicians. One's status as a physician can be very difficult to validate unless this information is cross-referenced against state licensing data. Physicians with verified accounts on Twitter (ie, accounts with blue badges) are the most reliable, since Twitter has a rigorous verification process that involves checking credible sources, such as health system websites and media sources, and using other means for verification [28]. Since activity trends were similar between the nonverified and verified accounts, we are reasonably confident that our approach produced a reliable cohort of physicians. Second, while Twitter's official verification process via the blue badge offers the most reliable method for validating physician accounts within pragmatic means, this approach has its limitations. Twitter has accidentally verified fake accounts, has temporarily halted verifications multiple times to improve their internal processes, and has overall been slow to verify physicians—an issue Twitter has recently addressed during the COVID-19 pandemic [37-40]. Third, it is very likely that our report underestimates the number of active US physicians on Twitter because some physicians may not indicate their status as a clinician. Although we explored various methodologies for matching Twitter accounts, such as using state licensing data, these approaches were not practical for this study, and they will be considered for future investigations.

### Future Directions

There are many opportunities to build on this study. We are planning future studies to focus on specific events, such as the COVID-19 pandemic. For example, by segmenting the physician tweets by state and correlating them with epidemiological data, we plan to study whether there was increased physician activity during pandemic surges. Similarly, machine learning and statistical analyses could be used to explore whether physicians' Twitter activity can be used as a predictor for public health alerts in the future. Physician activity could also be used as a potential real-time barometer for physician well-being at a regional or national level; by applying natural language processing techniques, such as sentiment analysis, such a data set could provide a baseline measurement of physician morale at any given point in time. We are exploring more in-depth analyses of Twitter bios; topic modeling; and the correlation of self-reported identities with validated physician registries, such as medical licensing boards. Finally, we feel that our methodology and this first comprehensive study of physician activity on Twitter offer a foundation to building more accurate and precise algorithms for identifying physician accounts nationally and globally. We also provide a more complete characterization of the social media activity of US physicians than what has been possible in the past.

### Conclusion

More physicians are using Twitter and covering a wider range of topics in their tweets. Twitter is entering a new phase of mainstream use among physicians, even though physician activity still represents a tiny fraction of total Twitter activity. These trends are evidence that Twitter can be a valuable source of data that can be used to understand social trends, interests, and the perceptions of the physician community. Unlike the more stringent settings of clinical practice, physicians are opening up on social media and showing a different side of themselves that is typically protected in the professional clinical setting. Physicians on Twitter may be offering a unique, more comprehensive view into the physician psyche, making Twitter an intriguing platform for further study and exploration.

---

## Conflicts of Interest

None declared.

---

Multimedia Appendix 1
Boolean logic for identifying physician accounts on Twitter.
[TXT File , 3 KB - jmir_v24i9e37752_app1.txt ]

Multimedia Appendix 2
Top 100 tweeted hashtags and themes between 2016 and 2020.
[PNG File , 62 KB - jmir_v24i9e37752_app2.png ]

Multimedia Appendix 3
Word clouds for hashtags from 2016 to 2020.
[PNG File , 231 KB - jmir_v24i9e37752_app3.png ]

## References

1. Sayce D. The number of tweets per day in 2020. David Sayce. URL: https://www.dsayce.com/social-media/tweets-day/ [accessed 2021-04-23]
2. Hasan M, Orgun MA, Schwitter R. A survey on real-time event detection from the Twitter data stream. J Inf Sci 2017 Mar 17;44(4):443-463. [doi: 10.1177/0165551517698564]
3. Earle PS, Bowden DC, Guy M. Twitter earthquake detection: earthquake monitoring in a social world. Annals of Geophysics 2011;54(6):708-715 [FREE Full text] [doi: 10.4401/ag-5364]
4. Sakaki T, Okazaki M, Matsuo Y. Tweet analysis for real-time event detection and earthquake reporting system development. IEEE Trans Knowl Data Eng 2013 Apr;25(4):919-931. [doi: 10.1109/tkde.2012.29]
5. Li R, Lei KH, Khadiwala R, Chang KCC. TEDAS: A Twitter-based event detection and analysis system. 2012 Presented at: 2012 IEEE 28th International Conference on Data Engineering; April 1-5, 2012; Arlington, VA, USA p. 1273-1276. [doi: 10.1109/icde.2012.125]
6. Wang X, Gerber MS, Brown DE. Automatic crime prediction using events extracted from Twitter posts. 2012 Presented at: SBP 2012: Social Computing, Behavioral-Cultural Modeling and Prediction; April 3-5, 2012; College Park, MD, USA p. 231-238. [doi: 10.1007/978-3-642-29047-3_28]

7. Daneshjou R, Adamson AS. Twitter journal clubs: Medical education in the era of social media. JAMA Dermatol 2020 Jul 01;156(7):729-730. [doi: 10.1001/jamadermatol.2020.0315] [Medline: 32186655]
8. Ghosh P, Schwartz G, Narouze S. Twitter as a powerful tool for communication between pain physicians during COVID-19 pandemic. Reg Anesth Pain Med 2021 Feb;46(2):187-188. [doi: 10.1136/rapm-2020-101530] [Medline: 32321859]
9. Schwenk ES, Jaremko KM, Park BH, Stiegler MA, Gamble JG, Chu LF, et al. I tweet, therefore I learn: An analysis of Twitter use across anesthesiology conferences. Anesth Analg 2020 Feb;130(2):333-340. [doi: 10.1213/ANE.0000000000004036] [Medline: 31124801]
10. Chisolm MS. Social media in medicine: The volume that Twitter built. Int Rev Psychiatry 2015 Apr;27(2):83-84. [doi: 10.3109/09540261.2015.1026239] [Medline: 25828741]
11. Chretien KC, Azar J, Kind T. Physicians on Twitter. JAMA 2011 Feb 09;305(6):566-568. [doi: 10.1001/jama.2011.68] [Medline: 21304081]
12. Riddell J, Brown A, Kovic I, Jauregui J. Who are the most influential emergency physicians on Twitter? West J Emerg Med 2017 Feb;18(2):281-287 [FREE Full text] [doi: 10.5811/westjem.2016.11.31299] [Medline: 28210365]
13. Sharp SP, Mackenzie DG, Ong DSY, Mountziaris PM, Logghe HJ, Ferrada P, et al. Factors influencing the dissemination of tweets at the American College of Surgeons Clinical Congress 2018. Am Surg 2021 Apr;87(4):520-526. [doi: 10.1177/0003134820950680] [Medline: 33106021]
14. @Bob_Wachter. 1/ Covid (@UCSF) Chronicles, Day 80 Local update: @UCSFHospitals, just 8 pts, 3 vented, no change from yesterday. In SF, cases still steady at ~20-30/d; and still no new deaths in a week (cumulative cases in Fig). 43 pts in hospitals, up 4 from yest; ICU pts 16, down from 19. 2020 Jun 05. URL: https://twitter.com/Bob_Wachter/status/1269096800096407552 [accessed 2022-06-05]
15. @Bob_Wachter. SF folks: it's now a big-time surge. No longer just cases (~500/d reported, so true # >2K, or ~250/100K/d). Also major uptick in hospitalizations: 98 in SF (vs 18 six wks ago). @UCSFHospitals 41 pts in hospital, twice April #. If you're trying to stay well, time to up your game. 2022 May 30. URL: https://twitter.com/Bob_Wachter/status/1531315837264482304 [accessed 2022-06-05]
16. @EricTopol. It's not over, by any means, but the chances of dying from Covid keep declining. 2022 Jun 05. URL: https://twitter.com/EricTopol/status/1533485528720416768 [accessed 2022-06-05]
17. @Craig_A_Spencer. Hospitalizations are increasing as Omicron cases reach record highs in NYC and all across the country. But as I noticed on my recent ER shifts, take a look at who needs admission to the hospital for severe Covid? TL;DR It's the unvaccinated. 2021 Dec 30. URL: https://twitter.com/Craig_A_Spencer/status/1476772672994488328 [accessed 2022-06-05]
18. Liu L, Woo BKP. Twitter as a mental health support system for students and professionals in the medical field. JMIR Med Educ 2021 Jan 19;7(1):e17598 [FREE Full text] [doi: 10.2196/17598] [Medline: 33464210]
19. Leibowitz MK, Scudder MR, McCabe M, Chan JL, Klein MR, Trueger NS, et al. Emergency medicine influencers' Twitter use during the COVID-19 pandemic: A mixed-methods analysis. West J Emerg Med 2021 Mar 22;22(3):710-718 [FREE Full text] [doi: 10.5811/westjem.2020.12.49213] [Medline: 34125051]
20. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. AMIA Annu Symp Proc 2014 Nov 14;2014:924-933 [FREE Full text] [Medline: 25954400]
21. Gabarron E, Dorronzoro E, Rivera-Romero O, Wynn R. Diabetes on Twitter: A sentiment analysis. J Diabetes Sci Technol 2019 May;13(3):439-444 [FREE Full text] [doi: 10.1177/1932296818811679] [Medline: 30453762]
22. Doogan C, Buntine W, Linger H, Brunt S. Public perceptions and attitudes toward COVID-19 nonpharmaceutical interventions across six countries: A topic modeling analysis of Twitter data. J Med Internet Res 2020 Sep 03;22(9):e21419 [FREE Full text] [doi: 10.2196/21419] [Medline: 32784190]
23. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
24. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
25. Cevik M, Ong DSY, Mackenzie G. How scientists and physicians use Twitter during a medical congress. Clin Microbiol Infect 2019 Dec;25(12):1561.e7-1561.e12 [FREE Full text] [doi: 10.1016/j.cmi.2019.04.030] [Medline: 31102781]
26. Margus C, Brown N, Hertelendy AJ, Safferman MR, Hart A, Ciottone GR. Emergency physician Twitter use in the COVID-19 pandemic as a potential predictor of impending surge: Retrospective observational study. J Med Internet Res 2021 Jul 14;23(7):e28615 [FREE Full text] [doi: 10.2196/28615] [Medline: 34081612]
27. Kilic Y, Chauhan D, Avery P, Horwood N, Nakov R, Disney B, et al. The public's attitude towards doctors' use of Twitter and perceived professionalism: an exploratory study. Clin Med (Lond) 2021 Sep;21(5):e475-e479 [FREE Full text] [doi: 10.7861/clinmed.2021-0357] [Medline: 34507932]
28. Twitter verification requirements - how to get the blue check. Twitter. URL: https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts [accessed 2021-06-28]
29. Urban Dictionary: TCOT. Urban Dictionary. URL: https://www.urbandictionary.com/define.php?term=TCOT [accessed 2021-07-04]

30. Benjamin Y. The secret Twitter war for America’s independents: #tcot vs #p2. SFGate. 2012 Nov 19. URL: https://web.archive.org/web/20121120014027/https://blog.sfgate.com/ybenjamin/2010/07/27/the-secret-twitter-war-for-americas-independents-tcot-vs-p2/ [accessed 2021-07-24]
31. Does my project need review by the IRB. UC Davis Office of Research. URL: https://research.ucdavis.edu/policiescompliance/irb-admin/researchers/irb-submissions/new-projects/do-i-need-irb-review/ [accessed 2022-08-11]
32. Rogers EM, Singhal A, Quinlan MM. Diffusion of innovations. In: Stacks DW, Salwen MB, editors. An Integrated Approach to Communication Theory and Research, Second Edition. New York, NY: Routledge; 2009.
33. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. Science 2019 Jan 25;363(6425):374-378. [doi: 10.1126/science.aau2706] [Medline: 30679368]
34. Osmundsen M, Bor A, Vahlstrup PB, Bechmann A, Petersen MB. Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. Am Polit Sci Rev 2021 May 06;115(3):999-1015. [doi: 10.1017/s0003055421000290]
35. Memon SA, Carley KM. Characterizing COVID-19 misinformation communities using a novel Twitter dataset. arXiv Preprint posted online on September 19, 2020 [FREE Full text] [doi: 10.48550/arXiv.2008.00791]
36. Huang B, Carley KM. Disinformation and misinformation on Twitter during the novel coronavirus outbreak. arXiv Preprint posted online on June 7, 2020 [FREE Full text] [doi: 10.48550/arXiv.2006.04278]
37. Porter J. Twitter admits it verified several fake accounts. The Verge. 2021 Jul 13. URL: https://www.theverge.com/2021/7/13/22575025/twitter-verified-fake-accounts-botnet-researcher [accessed 2022-06-05]
38. @verified. We’ve temporarily hit pause on rolling out access to apply for Verification so we can make improvements to the application and review process. For those who have been waiting, we know this may be disappointing. We want to get things right, and appreciate your patience. 2021 Aug 13. URL: https://twitter.com/verified/status/1426214732403466241 [accessed 2022-06-05]
39. Clark M. Twitter pauses its verification program rollout after giving fake accounts blue checks. The Verge. 2021 Aug 13. URL: https://www.theverge.com/2021/8/13/22623714/twitter-verification-program-pause-fake-accounts-review [accessed 2022-06-05]
40. Neporent L. Coronavirus social: Twitter hustles to verify physicians. Medscape. 2020 May 12. URL: http://www.medscape.com/viewarticle/930360 [accessed 2022-06-05]

*Edited by T Leung; submitted 05.03.22; peer-reviewed by J Luo, J Li, D Pessoa, S El kefi; comments to author 25.03.22; revised version received 05.06.22; accepted 12.07.22; published 06.09.22.*

*Please cite as:*
*Nakagawa K, Yang NT, Wilson M, Yellowlees P*
*Twitter Usage Among Physicians From 2016 to 2020: Algorithm Development and Longitudinal Analysis Study*
*J Med Internet Res 2022;24(9):e37752*
*URL: https://www.jmir.org/2022/9/e37752*
*doi:10.2196/37752*
*PMID:36066939*

Original Paper

# Public Discourse and Sentiment Toward Dementia on Chinese Social Media: Machine Learning Analysis of Weibo Posts

Dexia Kong[1], MSW, MBE, PhD; Anfan Chen[2], PhD; Jingwen Zhang[3,4], PhD; Xiaoling Xiang[5], PhD; W Q Vivian Lou[6], MSW, PhD; Timothy Kwok[7], MD; Bei Wu[8], PhD

[1]Department of Social Work, Chinese University of Hong Kong, Hong Kong, Hong Kong
[2]School of Journalism and Communication, The Chinese University of Hong Kong, Hong Kong, Hong Kong
[3]Department of Communication, University of California, Davis, Davis, CA, United States
[4]Department of Public Health Sciences, University of California, Davis, Davis, CA, United States
[5]School of Social Work, University of Michigan, Ann Arbor, MI, United States
[6]Department of Social Work & Social Administration, Sau Po Centre on Ageing, University of Hong Kong, Hong Kong, Hong Kong
[7]Faculty of Medicine, Chinese University of Hong Kong, Hong Kong, Hong Kong
[8]Rory Meyers College of Nursing, New York University, New York, NY, United States

**Corresponding Author:**
Anfan Chen, PhD
School of Journalism and Communication
The Chinese University of Hong Kong
NAH313B, Humanities Building
New Asia College
Hong Kong, 999077
Hong Kong
Phone: 852 68436916
Email: anfanchen@cuhk.edu.hk

## Abstract

**Background:** Dementia is a global public health priority due to rapid growth of the aging population. As China has the world's largest population with dementia, this debilitating disease has created tremendous challenges for older adults, family caregivers, and health care systems on the mainland nationwide. However, public awareness and knowledge of the disease remain limited in Chinese society.

**Objective:** This study examines online public discourse and sentiment toward dementia among the Chinese public on a leading Chinese social media platform Weibo. Specifically, this study aims to (1) assess and examine public discourse and sentiment toward dementia among the Chinese public, (2) determine the extent to which dementia-related discourse and sentiment vary among different user groups (ie, government, journalists/news media, scientists/experts, and the general public), and (3) characterize temporal trends in public discourse and sentiment toward dementia among different user groups in China over the past decade.

**Methods:** In total, 983,039 original dementia-related posts published by 347,599 unique users between 2010 and 2021, together with their user information, were analyzed. Machine learning analytical techniques, including topic modeling, sentiment analysis, and semantic network analyses, were used to identify salient themes/topics and their variations across different user groups (ie, government, journalists/news media, scientists/experts, and the general public).

**Results:** Topic modeling results revealed that symptoms, prevention, and social support are the most prevalent dementia-related themes on Weibo. Posts about dementia policy/advocacy have been increasing in volume since 2018. Raising awareness is the least discussed topic over time. Sentiment analysis indicated that Weibo users generally attach negative attitudes/emotions to dementia, with the general public holding a more negative attitude than other user groups.

**Conclusions:** Overall, dementia has received greater public attention on social media since 2018. In particular, discussions related to dementia advocacy and policy are gaining momentum in China. However, disparaging language is still used to describe dementia in China; therefore, a nationwide initiative is needed to alter the public discourse on dementia. The results contribute to previous research by providing a macrolevel understanding of the Chinese public's discourse and attitudes toward dementia, which is essential for building national education and policy initiatives to create a dementia-friendly society. Our findings indicate that dementia is associated with negative sentiments, and symptoms and prevention dominate public discourse. The development of strategies to address unfavorable perceptions of dementia requires policy and public health attention. The results further reveal that an urgent need exists to increase public knowledge about dementia. Social media platforms potentially could be leveraged for future dementia education interventions to increase dementia awareness and promote positive attitudes.

***(J Med Internet Res 2022;24(9):e39805)*** doi:10.2196/39805

**KEYWORDS**

dementia; public discourse; sentiment; Weibo; social media; machine learning; infodemiology; aging; elderly population; content analysis; topic modeling; thematic analysis; social support; sentiment analysis; public discourse

## Introduction

China has the largest population with dementia in the world. An estimated 9.5 million Chinese adults aged 60 years and older had dementia in 2015, accounting for roughly a quarter of the entire population with dementia worldwide [1,2]. Due to China's aging population, the number of individuals with dementia is projected to reach 23.3 million by 2030 [3]. However, previous studies have found that in China, only around 10% of older people with dementia have a clinical diagnosis, leaving the care needs of a substantial number of families impacted by dementia invisible [4]. Even in the metropolitan regions of China where dementia-related health care services and workforce are available, individuals wait 2 years on average between the onset of symptoms and the first medical visit [5,6]. Such delays are likely more common in underresourced regions and rural China. Furthermore, receiving a clinical diagnosis does not often lead to adequate care. One study reported that over 90% of patients with a dementia diagnosis remain inadequately managed due to a lack of training and knowledge among health care professionals [4,7]. Delayed diagnosis and care impose unnecessary burdens on individuals and family caregivers, as well as increase health care expenditures [8,9]. Extensive evidence indicates that limited knowledge of dementia is a significant predictor of delays in seeking dementia diagnosis and treatment [6,10,11]. An online survey of 10,562 Chinese individuals found that just half of the sample was aware of dementia risk factors [12].

Therefore, improving the public awareness of dementia has been identified as a significant global public health priority. Developing effective approaches to raising the public awareness of dementia requires a solid understanding of current public discourse and attitudes toward the the disease to identify knowledge gaps and priority areas for intervention. However, previous studies in China so far have relied on regional samples or specific stakeholder groups (eg, older adults and their family caregivers [13], mental health providers [14], and health care professionals [15]), with most studies using small samples, thereby providing limited generalizability. As a result, our understanding of dementia awareness, sense making, and comprehension among the Chinese population remains limited.

The phenomenal growth in social media users has provided new avenues for public health research and strategies for health promotion [16]. With 940 million internet users, China reached an internet penetration rate of 67% in 2020 [17]. Obtaining health-related information from social media has become a common practice in China [18]. Not only have platforms such as blogs, online forums, and social networking sites increased patients' access to information and support, they also have offered a virtual arena in which the general public can express views, opinions, and sentiments about various health conditions. A previous study found that social media has provided a primary avenue through which Chinese people freely share, receive, and circulate knowledge about dementia [19]. Analyzing social media data is a cost-effective and unobtrusive way to assess public perceptions and attitudes toward dementia for 2 reasons. First, compared with survey research, naturally created online user-generated data on social media are less vulnerable to social desirability and recall bias [20]. Using computational tools to analyze social media data concerning dementia may reveal additional information that epidemiological survey research fails to uncover. For instance, social media provides unique sources of information that might be challenging to measure in conventional public health research (eg, personal opinions and experiences), thereby offering new avenues for population health research [21]. Second, utilizing social media data also allows for a more complete understanding of the public's knowledge and attitudes than previous studies with regional samples, given the high penetration of social media use. Machine learning of user-generated data could identify macrolevel social determinants of population health, which may offer critical insights for public health research, policy, and practice [20].

Weibo (Sina Corporation), founded in 2009, is a Twitter-esque microblogging platform in China. Weibo has 249 million daily active users posting images and short personal stories and perspectives of up to 140 Chinese characters per post. Weibo has grown to be 1 of China's most prevalent social media platforms. Public opinions expressed on Weibo often reflect current social concerns/sentiments and policy issues in China [22]. Weibo users can follow others, be followed, and create/share/receive information about social events (eg, posting/reposting others' posts on homepages and broadcasting personal posts to followers), and they can attach photos, videos, universal resource locators (URLs), and emojis to posts [22]. Weibo offers individual (eg, members of the general population, celebrities, and government officials) and organizational accounts that share various information and opinions.

This study analyzed longitudinal user-generated microblogging data from Weibo between January 1, 2010, and July 31, 2021, using machine learning techniques. The study aimed to (1) assess and examine public discourse and sentiment toward dementia among the Chinese public, (2) determine the extent to which dementia-related discourse and sentiment vary among different user groups (ie, government, journalists/news media, scientists/experts, and the general public), and (3) characterize temporal trends in public discourse and sentiment toward dementia among different user groups during the past decade. To the best of our knowledge, this study is the first to examine the public discourse on dementia among Chinese social media users. Findings from this study could inform the development of public health campaigns and interventions aimed at increasing dementia awareness and subsequently improve disease detection and management. Moreover, this study's findings will contribute to concerted efforts to respond to the World Health Organization's (WHO) global action plan to raise dementia awareness.

## Methods

### Data Collection and Preprocessing

We used the keyword-based advanced search function provided by Weibo to retrieve relevant data. The selected keywords were informed by sources that included existing studies on media/social media discussions of Chinese dementia/Alzheimer disease [23] and a careful review of 294 Weibo forums (#daily routine of old age idiotic# and #age idiotic# among others) and other social media forums (ie, Baidu Tieba, Zhihu) concerning dementia and Alzheimer disease by 2 authors. Combinations of dementia-related keywords, including "senile dementia" ("老年痴呆" in Chinese), "dementia" ("失智症"), "Alzheimer" ("阿兹海默症"), "brain atrophy" ("脑萎缩"), and their variations (eg, "痴呆症," "认知障碍," "小脑萎缩," "脑梗," "失智症," "脑退化," "阿尔茨海默," "认知功能障碍"), were selected to retrieve relevant posts between January 1, 2010 (4 months after Weibo was launched in China), and July 20, 2021. After removing duplicates, 983,039 original posts (posts that started/initiated the thread, also known as thread posts; see Multimedia Appendix 1, Figure S1) published by 347,599 unique Weibo users were obtained for analysis. Furthermore, user information (eg, screen name, introduction, user-type status, verification status, follower number, following number, geolocation, education, marital status, gender, registration date, and other information) was also retrieved. Before the subsequent data analysis procedures occurred (ie, latent Dirichlet allocation [LDA] topic modeling, sentiment analysis, and semantic network analyses), we conducted text preprocessing [24] to clean the data (see Multimedia Appendix 1, "Text Preprocessing Procedures," for detailed preprocessing procedures).

### Data Analysis

#### *Dementia-Related Themes*

We conducted topic modeling using LDA to identify dementia-related themes. Topic modeling is a computer-assisted content analysis machine learning technique that is semiautomated and unsupervised [25-29]. It is a form of semantic analysis in which statistical algorithms are used to identify abstract or hidden themes that arise from a large corpus of text. This approach presumes that each collection of documents has a given number of hidden themes and that terms associated with a certain subject are used commonly in close proximity to one another. LDA, 1 of the most widely used topic-modeling techniques, classifies texts into latent topics, each of which is represented by an extracted cluster of keywords based on the computed probability of keyword co-occurrence [30].

Based on the results of the LDA modeling process, we chose the optimal and highly granular k value of 30 topics (for details, see Multimedia Appendix 1, "Interpreted Topics With Top 10 Words and Associated Frames"). Specifically, we used statistical indicators of model fit to determine the optimal number of topics, during which a range of models containing 10-300 topics were estimated. Models were examined using skips of 10 for 10-100 topics, skips of 20 for 101-200 topics, and skips of 50 for 201-300 topics. Next, a smaller range of k-topics was selected based on 4 different model fit indicators. We chose the k areas in which indicators of accuracy [31] and density [32] were minimized, and indicators from latent concept modeling and a Bayesian Markov chain Monte Carlo algorithm were maximized [33,34]. For the chosen limited range, we reran the models without skips (eg, after the first iteration pointed to the range of 30-40, we modeled each of the possible 11 models for this corpus). Next, we examined 3 types of information to interpret and label topics and extract potential themes among them, including words with the highest loading on each topic, words that were both prevalent in and exclusive to each topic, and full Weibo posts that were most representative of each topic. Four coauthors, including a public health communication specialist, read and labeled the topics. Next, 3 coauthors independently extracted themes using open coding. Preliminary themes were discussed, and a final set of themes was reached via consensus building. This approach is consistent with the assumption that LDA-generated subjects may be recoded or regrouped into meaningful, condensed frames based on conceptual similarity [35].

#### *User Groups*

Users were classified into 1 of 4 types based on user information supplied and verified by Weibo (see Table 1 for a detailed description of the user groups): (1) government, including all levels and departments of the Chinese government; (2) journalists/news media, including newspapers, broadcasting companies, and news websites, as well as their affiliated personnel (eg, journalists, editors, and hosts/hostesses); (3) scientists/experts, including influential individuals and organizations in bioscience, medicine, neuroscience, and related fields (eg, scientists/researchers, research departments, and nongovernment organizations in the related field); and (4) users in the general public who were not included in the first 3 categories. Due to the Weibo data set's longitudinal nature (2010-2021), we performed analyses to determine temporal changes in public discourse and sentiment across different user groups.

**Table 1.** User group typologies.

| User type | Description | Examples | Users (N=347,599), n (%) | Posts (N=983,039), n (%) |
|---|---|---|---|---|
| Government | Different levels and departments of the Chinese government | Gansu Higher People's Court, the Publicity Department of the Central Committee of the Communist Party of China (CPC), the Tianjin Public Security Bureau | 11,829 (3.40) | 70,171 (7.14) |
| Scientists/experts | Users with expertise in neuroscience, including scientists/researchers in the related field, health care providers, and medical experts | Cao Pusheng, director of oncology and surgeon at the Nanning Hospital of Integrated Traditional Chinese and Western Medicine, and Yang Jian, chief physician of the Department of Neurology, Capital Institute of Pediatrics | 24,601 (7.08) | 140,924 (14.34) |
| Journalists/news Media | Journalists and news media, including newspapers, broadcasters, and news websites | *Beijing Youth Daily*, *National Business Daily* | 12,138 (3.49) | 92,030 (9.36) |
| General public | Users outside the first 3 groups or without specific identities | Lin Yinwei | 299,031 (86.03) | 679,914 (69.16) |

### Semantic Analysis

We conducted semantic network analysis to cross-validate the results from topic modeling. Specifically, public perceptions of dementia among Chinese netizens were characterized by analyzing the semantic networks of dementia discussions on Weibo among various user groups (eg, government, scientists/experts, journalists/news media, and the general public). Following previous studies' practices [36-38], this study adopted the semantic network analysis method, an analytical approach for deriving meaning from text. Microblogs were analyzed to identify grammatical structures and relationships between individual words [37]. The semantic networks of dementia discussions on Weibo among different types of users (ie, government, scientists/experts, journalists/news media, and the general public) were analyzed using the ForceAtlas2 layout configuration embedded in Gephi. The top 100 words by frequency in each corpus were retained for concise visualization and comprehension. Keywords used to search for data were removed from the final networks because predominant words are highly likely to link all the other words, which may distort the results [36].

### Sentiment Analysis

Sentiment analysis was conducted to ascertain the positive/negative emotions or feelings conveyed in the dementia-related posts. Sentiments were analyzed by using 2 emotion dictionaries (the *Chinese DLUT-Emotion Ontology* developed by the Dalian University of Technology's Information Retrieval Laboratory and the *Chinese Emotion Valence Dictionary*) in conjunction with dictionary-based techniques [39-41]. For each post, the algorithm creates a continuous sentiment score ranging from negative to positive, with the number's magnitude indicating sentiment strength. We further assessed whether each sentiment was associated with user types and post characteristics (eg, publication time and dementia-related themes) in regression models.

## Results

### Top Words in Public Discourse Concerning Dementia

Figure 1 displays the occurrence of each search term over 12 years. Some keywords, including "痴呆" ("stupid" or "idiotic") and "Alzheimer" ("阿尔兹海默"), occurred more often than others, such as "Alzheimer" ("阿兹海默"), "brain atrophy" ("脑萎缩"), "cognitive impairment" ("认知障碍"), "cerebral atrophy" ("小脑萎缩"), "cerebral infarction" ("脑梗"), "dementia" ("失智症"), "Alzheimer disease" ("脑退化"), and "cognitive dysfunction" ("认知功能障碍"). Figure 2 depicts the temporal distribution of dementia-related posts on Weibo from 2010 to 2021. Since 2018, dementia posts increased significantly.

**Figure 1.** Temporal distribution of search terms in collected posts. Note: “阿兹海默” and “阿尔兹海默” are 2 terms for Alzheimer disease used in simplified Chinese. “脑萎缩,” “认知障碍,” “小脑萎缩,” “脑梗,” “失智症,” “脑退化,” and “认知功能障碍” are synonyms/variations of Alzheimer disease, and 2 terms, including “老年痴呆” (“old age idiotic”) and “痴呆症” (“idiotic disease”) were merged into “痴呆” to display.

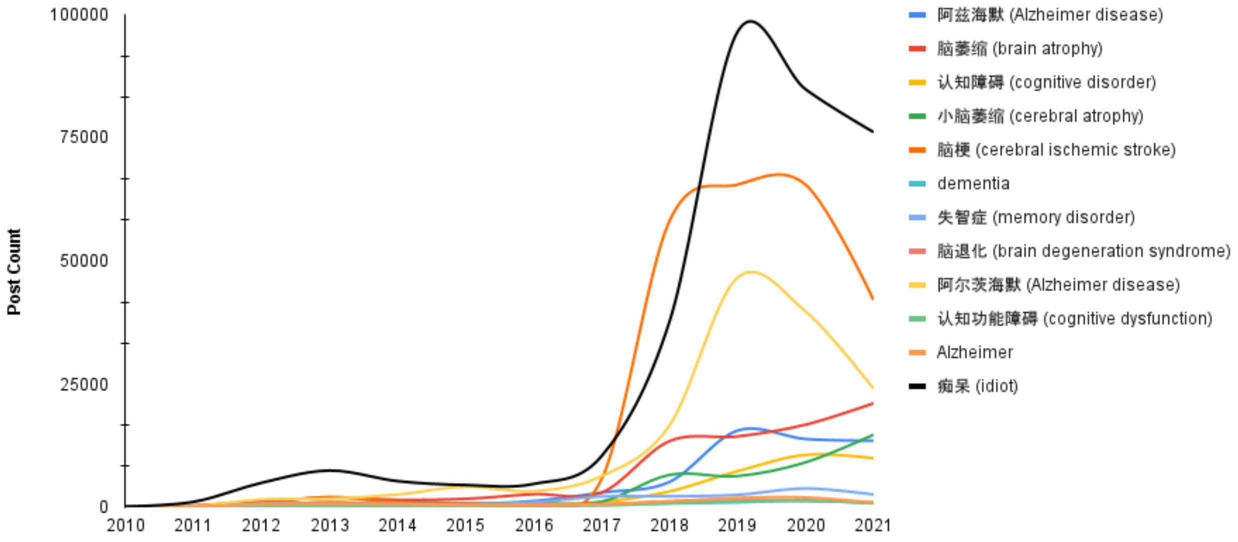
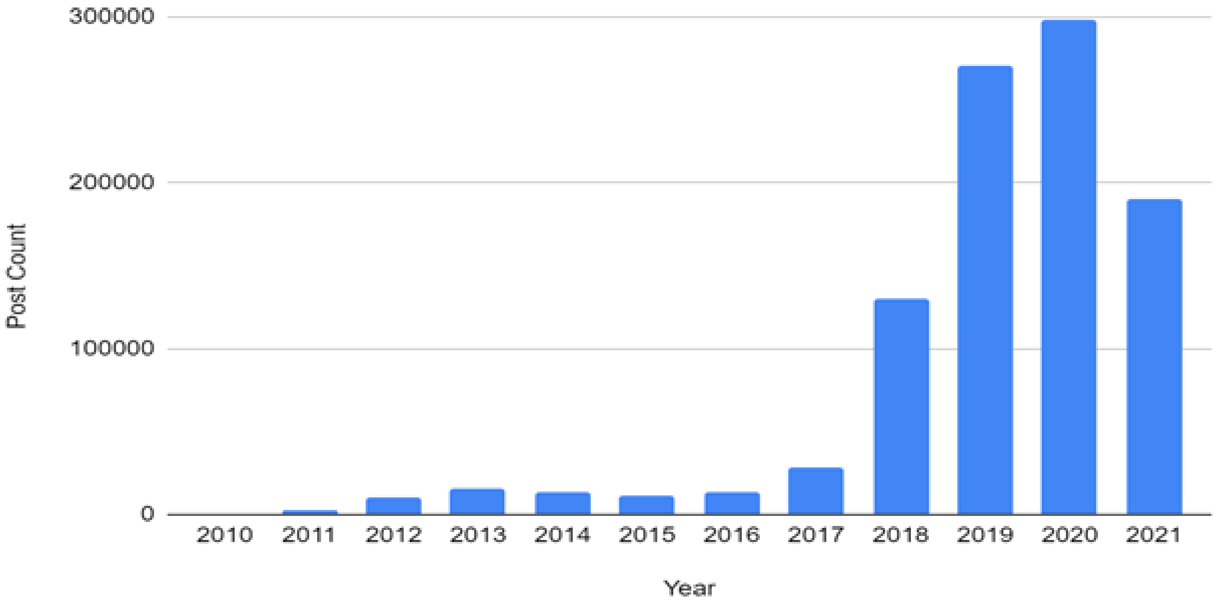


**Figure 2.** Temporal distribution of dementia-related posts on Weibo. Note: As a result of the sudden drop in data volume in 2021, the year’s data only cover January 1-July 17.



## Dementia Topics and Themes

We derived 30 latent topics related to dementia from the posts by evaluating the 983,039 dementia-related Weibo posts and reviewing prior research on social media discussions of dementia [42-46]. Table 2 provides the 30 topics and top 10 unique keywords for each topic (a native Chinese-speaking researcher translated the words into English). As illustrated in Table 2, 8 main dementia-related themes were identified, namely social support (n=143,505, 14.60%), advocacy and policy (n=55,905, 5.69%), prevention (n=213,453, 21.71%), media coverage (n=114,949, 11.69%), symptoms (n=246,124, 25.04%), treatment (n=90,959, 9.25%), personal experience (n=87,012, 8.85%), and raising awareness (n=31,132, 3.17%); see Multimedia Appendix 1, “Interpreted Topics With Top 10 Words and Associated Frames,” for frame examples.

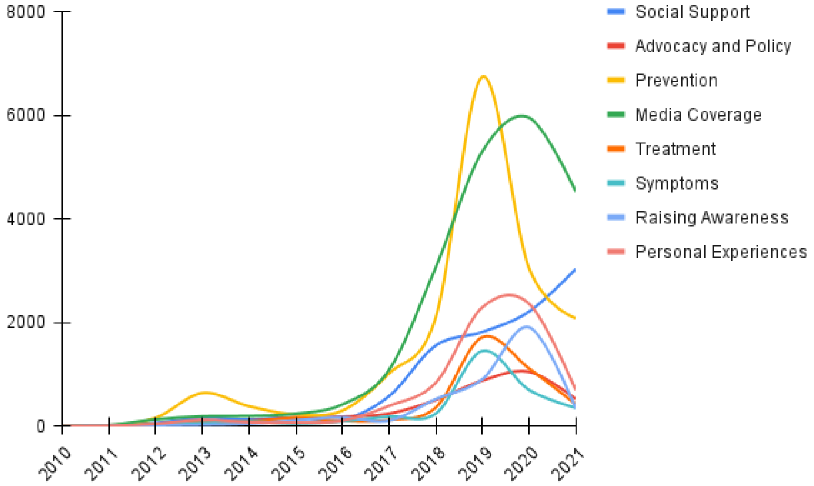
**Table 2.** Frames identified from the 30 dementia-related topics on Weibo.

| Frame | Posts (N=983,039), n (%) | Description | Subtopics | Representative keywords |
|---|---|---|---|---|
| Social support | 143,505 (14.60) | Support received from formal and informal sources | 0, 7, and 10 | mom, grandma, take care, children, grandpa, hospital, living, dad, go home, hospitalization, parents, passed away, father, repost/forward, good men/good Samaritan, hope, love, mother, Waterdrop Inc., help, please, life, family, home, elderly/older people, suffering |
| Advocacy and policy | 55,905 (5.69) | Discussions related to policy issues/changes and advocacy initiatives | 1, 2, and 21 | China, work, overturn, Bailao Yang, human right, street, income, job, service, project, activity, community, society, center, provide, concern, country, medical treatment, free, people with disability, hard-working, temporary worker, Disabled Persons Federation, civil affairs, alone, dismissal, homeless |
| Prevention | 213,453 (21.71) | Preventive biological and lifestyle behaviors | 3, 5, 8, 16, 18, 20, 23, 26, and 27 | prevention, resolve toxin, brain cells, weight loss, detoxify, suppression, antiaging, protection, effect, pitaya, memory, relaxing bowels, risk, increase, incur, cause, hospitalization, decrease, food, health, cardiovascular, senescence, disease, strengthen, chocolate, emotion, exercise, improve, diabetes, efficacy, enhance, immunity, contain, prevention, stay up all night, prevent, live to the old age, exercise/movement, training, remind, share, expert, blood vessels, sleep, diet, function, reduce, senile dementia/Alzheimer disease, long-term |
| Media coverage | 114,949 (11.69) | Description of dementia on media outlets | 4, 9, 13, and 24 | news, Weibo, video, driver, recent, doctor, life, bus, send to hospital, one, passenger, an older person, patient, situation, emergency, older persons, being lost, find, older spouse, repost/forward, family, go home, search, tracing notice, phone, search people, help, wear, height, older people, suffering, policeman, one, family members, senile dementia, found, report to police, police station, son, home, be informed of, getting lost, gender, movie, time, father, story, world, trapped, U.S., diagnosis, Alzheimer disease, pass away |
| Symptoms | 246,124 (25.04) | Discussion of different manifestations of dementia | 11, 14, 15, 17, 22, and 29 | forget, remember, brain, thing, memory, suspicion, ability to remember, a little, omen, patients, cause, diagnosis, decrease, function, usually, myocardial infarction, blood vessel, high blood pressure, stroke, disease, thrombosis, cerebral hemorrhage, blood pressure, heart disease, diabetes, garbage, fat, cardiovascular, blood lipids, find, morning, feeling, come back, go home, go out, forget, symptoms, result in, cognition, performance, influence, neuro, decline, atrophy, dizziness, insomnia, cerebral thrombosis, spinal disease, headache, hand numbness, head, qi and blood, cervical spine, meridian, neck, relief, comfort, nausea |
| Treatment | 90,959 (9.25) | Potential treatment options | 6 and 19 | treatment, brain atrophy, patient, symptom, cerebral infarction, hospital, doctor, check, disorder, walk, limbs, state of illness, condition, rehabilitation, recovery, traditional Chinese medicine, body, talk, drug, disease, global clinical, protein, neuro, cognitive, new drug |
| Personal experience | 87,012 (8.85) | Sharing individual/family experiences with dementia | 12 and 28 | older people, suffering, forget, memory, grandma, remember, restaurant, unforgettable, daughter, mother, grandpa, young, cognition, life, barrier, time, one type, cognitive impairment/disorder, understanding, emotion, ability, world, society, self, think, things, change, like, way, learn |
| Raising awareness | 31,132 (3.17) | Posts focusing on fostering greater understanding, attention, and positive attitudes toward dementia | 25 | Alzheimer disease, patient, World Alzheimer's Day, China, older people, health, memory, dementia, elderly, prevention, senile, cognition, disease, our country, treatment |

## Temporal Distribution of Frames

As shown in Figure 3, symptoms, prevention, and social support are the most prevalent dementia frames, while advocacy and policy have been gaining obvious prominence since 2019. Posts about raising awareness of dementia have consistently remained the least used dementia theme in the past decade.

**Figure 3.** Temporal distribution of dementia-related post frames.



## Theme–User Group Analysis

To examine possible differences in dementia frames across different user groups, we first conducted an 8 (frames) × 4 (user groups) chi-square test, which revealed that dementia frame adoptions significantly differed across the 4 types of users ($\chi^2_{21}$=147,217.420, *P*<.001). Specifically, as shown in Table 3, prevention (n=46,437, 32.95%), symptoms (n=30,238, 21.46%), and treatment (n=31,137, 22.09%) frames were relatively salient, with a high proportion for scientists/experts. With regard to the government, media coverage (n=21,133, 30.12%), prevention (n=16,707, 23.81%), and social support (n=9853, 14.04%) were the most commonly used frames. For journalists/news media users, media coverage (n=23,041, 25.04%), prevention (n=19,264, 20.93%), and social support (n=16,267, 17.67) ranked much higher than the other 5 dementia frames. With regard to the general public, treatment (n=216,321, 31.82%), prevention (n=131,978, 19.41%), and social support (n=111,431, 16.39%) were the most frequently adopted frames, followed by media coverage (n=65,167, 9.73%), personal experience (n=62,865, 9.58%), symptoms (n=45,208, 6.76%), advocacy/policy (n=43,369, 6.78%), and raising awareness (n=18, 0.003%). Overall, for dementia framing, the general public is more aligned with scientists/experts, while the government and journalists/news media users share a similarity.

**Table 3.** Differences in dementia frames among different user groups.

| Function | Government (n=70,171), n (%) | Scientists/experts (n=140,924), n (%) | Journalists/newsmedia (n=92,030), n (%) | General public (n=679,914), n (%) |
|---|---|---|---|---|
| Social support | 9853 (14.04) | 6494 (4.61) | 16,267 (17.67) | 111,431 (16.39) |
| Advocacy and policy | 3788 (5.40) | 4233 (3.00) | 4250 (4.62) | 46,132 (6.78) |
| Prevention | 16,707 (23.81) | 46,437 (32.95) | 19,264 (20.93) | 131,978 (19.41) |
| Media coverage | 21,133 (30.12) | 4514 (3.20) | 23,041 (25.04) | 65,167 (9.58) |
| Symptoms | 4206 (6.00) | 30,238 (21.46) | 10,303 (11.20) | 45,988 (6.76) |
| Treatment | 3285 (4.68) | 31,137 (22.09) | 5909 (6.42) | 216,321 (31.82) |
| Personal experience | 7020 (10.00) | 7140 (5.07) | 8888 (9.66) | 62,865 (9.25) |
| Raising awareness | 4179 (5.95) | 10,731 (7.61) | 4108 (4.46) | 32 (0.003) |

## Themes in Public Discourse Concerning Dementia

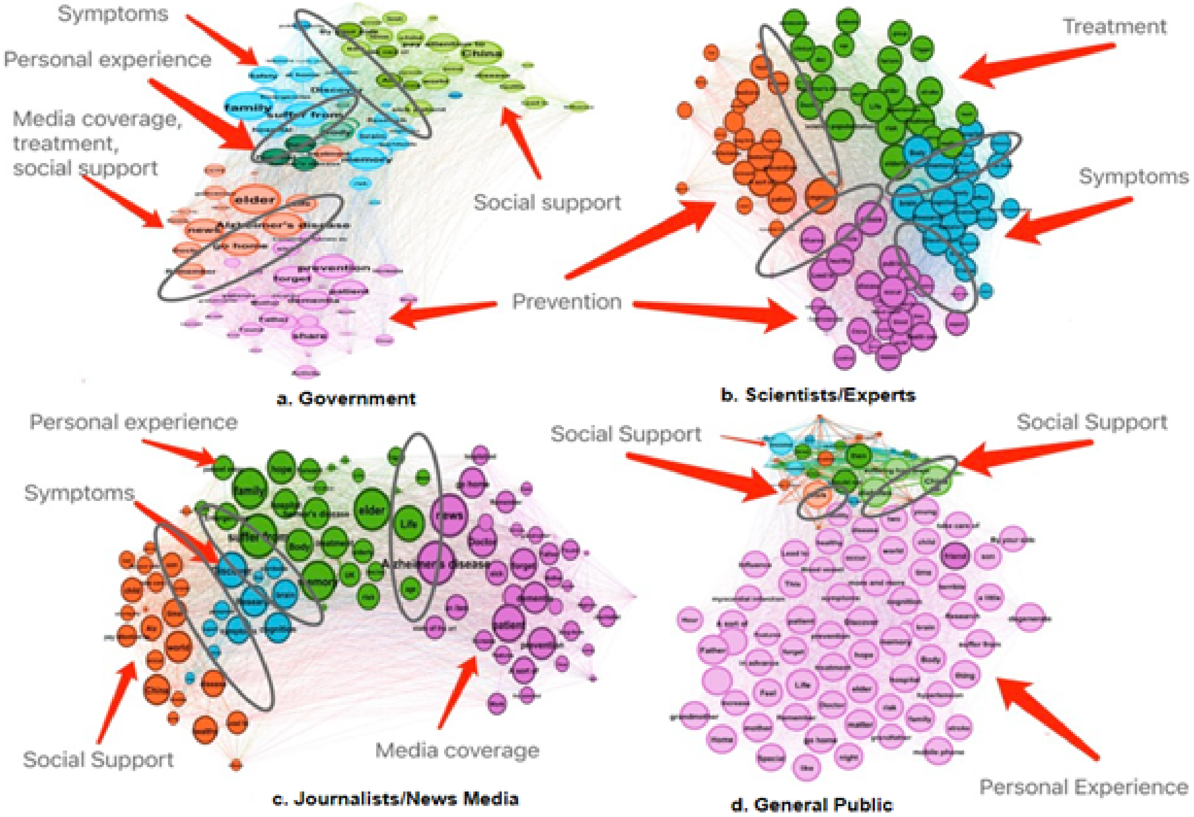
Table 4 presents the clusters and top word associations, as well as the percentage share of the network for each subcluster. We identified 5 word clusters in the government semantic network. Among them, cluster 3 (33.33% of the semantic network) was the most prevalent frame, focusing on prevention, followed by cluster 2 (25.00%) on social support; cluster 5 (20.83%) on symptoms; cluster 1 (14.58%) on media coverage, treatment, and social support; and cluster 4 (6.25%) on personal experience. For scientists/experts, 4 clusters of words were identified: cluster 2 (30.53%) and cluster 3 (23.16%) focusing mostly on prevention, cluster 4 (24.21%) on treatment, and cluster 1 (22.11%) on symptoms. For journalists/news media, media coverage emerged as the most prevalent word cluster (37.76%), followed by personal experience (ie, cluster 2: 27.55%), social support (ie, cluster 4: 23.47%), and symptoms (ie, cluster 1: 11.22%). In terms of the general public's semantic network, personal experience (ie, cluster 4: 64.65% of the semantic network) was the dominant cluster, followed by social support (ie, clusters 1, 2, and 3: 10.10%, 16.16%, and 9.1% of the semantic network, respectively).

As shown in Figure 4, the distance between 2 clusters indicates their coshared words' proportion and their association in the semantic network. For the government, prevention was closer to media coverage, treatment, and social support, although symptoms shared more keywords with personal experience and social support. For scientists/experts, prevention had more connections with treatment and symptoms. For journalists/news media, personal experiences were related more to media coverage and symptoms, while symptoms also shared more similarities with social support. For the general public, the most prominent cluster association was between personal experiences and social support.

**Table 4.** Summary results of network modularity analysis of semantic networks by 4 user groups.

| Network and clusters | Associated frames | Top terms | Share of network, (%) |
|---|---|---|---|
| **Government** | | | |
| 1 | Media coverage, treatment, social support | CCTV, elder, news, go home, life, doctor, remember, treatment, policeman, ask for help | 14.58% |
| 2 | Social support | pay attention to, son, disease, child, by your side, take care of, lead to, healthy, parents, spouse | 25.00% |
| 3 | Prevention | prevention, forget, sick, father, mother, suffer from, patience, activity, life, increase | 33.33% |
| 4 | Personal experience | disease, hope, body, children, chronic, forward | 6.25% |
| 5 | Symptoms | suffer from, memory, discover, hospital, brain, safety, research, emergencies, risk, cognition, personal | 20.83% |
| **Scientists/experts** | | | |
| 1 | Symptoms | brain, memory, research, body, discover, cognition, early, symptoms, hospital, effects, dysfunction | 22.11% |
| 2 | Prevention | reduce, publish, lead to, healthy, article, health care, blood, drug, expert, control, cardiovascular, blood vessel, diabetes | 30.53% |
| 3 | Prevention | improve, prevention, patient, medicine, food, performance, increase, features, long-term | 23.16% |
| 4 | Treatment | elderly, risk, treatment, life, science popularization, doctor, stroke, factors, hypertension, diet | 24.21% |
| **Journalists/news media** | | | |
| 1 | Symptoms | discover, brain, cognition, personnel, sleep, worldwide, publish, symptoms, expert, drug | 11.22% |
| 2 | Personal experience | suffer from, elder, family, memory, life, hope, body, treatment, hospital, emergencies, risk | 27.55% |
| 3 | Media coverage | news, patient, doctor, prevention, forget, go home, sick, state-of-art, father, remember, work, increase | 37.76% |
| 4 | Social support | work, healthy, China, time, disease, child, son, lead to, pay attention to, occur, take care of, by your side | 23.47% |
| **General public** | | | |
| 1 | Social support | work, disabled, disabled people’s federation, child union, China Unicom, sweat, more than 30 years, repay debt, get back, flaunt, Kuomintang | 10.10% |
| 2 | Social support | diabetes, China, suffer from, street, since ancient times, overthrow, temporary worker, superiority, homeless, dismissal, kill for life | 16.16% |
| 3 | Social support | income, nowadays, human rights, alone, civil affair, hard-earned money, tens of millions, fabricated | 9.1% |
| 4 | Personal experience | life, father, matter, elder, memory, mother, discover, degenerate, home, life, hope, mother | 64.65% |

**Figure 4.** Semantic network visualization of dementia-related discussions among different users on Weibo. Note: All words are in lowercase.



## Sentiment Toward Dementia

Results from the sentiment analyses are presented in Table 5. The findings revealed that dementia-related Weibo posts published in 2011 (B=2.85, *P*=.002), 2012 (B=2.64, *P*=.002), and 2016 (B=1.69, *P*=.05) were significantly more positive compared with those published in 2010 (see Multimedia Appendix 1, “Sentiment Analysis Results by Year, Frame, and User Type,” for sentiment mean values for each year). In terms of user groups, government users (B=5.81, *P*<.001), scientists/experts (B=3.22, *P*<.001), and journalists/news media (B=5.11, *P*<.001) tended to be more positive than the general public concerning dementia discourse on Weibo. Moreover, when compared to personal experiences, only posts on advocacy and policy were associated with positive sentiments (B=3.16, *P*<.001), whereas posts on social support (B=–5.27, *P*<.001), prevention (B=–3.17, *P*<.001), media coverage (B=–3.52, *P*<.001), symptoms (B=–3.55, *P*<.001), treatment (B=–9.49, *P*<.001), and raising awareness (B=–4.34, *P*<.001) tended to express more negative emotions.

**Table 5.** Regression model predicting sentiment with dementia-related messages on Weibo ($R^2$=6%, N=983,039).

| Year, frame, and user type | Estimate | 95% CI | *P* value[a] |
|---|---|---|---|
| **Dementia-related frames (reference group=personal experience)** | | | |
| Social support | –5.27 | –5.42 to –5.12 | *<.001* [b] |
| Advocacy and policy | 3.16 | 2.97-3.35 | *<.001* [b] |
| Prevention | –3.17 | –3.31 to –3.04 | *<.001* [b] |
| Media coverage | –3.52 | –3.67 to –3.36 | *<.001* [b] |
| Symptoms | –3.55 | –3.71 to –3.38 | *<.001* [b] |
| Treatment | –9.49 | –9.63 to –9.36 | *<.001* [b] |
| Raising awareness | –4.34 | –4.57 to –4.11 | *<.001* [b] |
| **User groups (reference group=general public)** | | | |
| Government | 5.81 | 5.67-5.95 | *<.001* [b] |
| Scientists/experts | 3.22 | 3.11-3.32 | *<.001* [b] |
| Journalists/news media | 5.11 | 4.98-5.23 | *<.001* [b] |
| **Year (reference group=2010)** | | | |
| 2011 | 2.85 | 1.04-4.66 | *.002* [c] |
| 2012 | 2.64 | 0.94-4.34 | *.002* [c] |
| 2013 | 1.56 | –0.13 to 3.24 | .07 |
| 2014 | 0.54 | –1.15 to 2.23 | .53 |
| 2015 | 1.29 | –0.40 to 2.99 | .14 |
| 2016 | 1.69 | 0.00-3.38 | *.05* [d] |
| 2017 | 0.51 | –1.17 to 2.18 | .55 |
| 2018 | 1.42 | –0.24 to 3.09 | .09 |
| 2019 | 0.48 | –1.18 to 2.15 | .57 |
| 2020 | –0.46 | –2.13 to 1.20 | .59 |
| 2021 | 0.52 | –1.15 to 2.18 | .54 |
| Constant | 7.63 | 5.96-9.30 | .01[d] |

[a] Italicized *P* values are significant.

[b] *P*<.001.

[c] *P*<.01.

[d] *P*<.05.

## Discussion

### Principal Findings

By analyzing 983,039 Weibo posts from January 1, 2010, to July 20, 2021, using machine learning techniques, this study discovered prominent themes and feelings tied to dementia among the Chinese population, as well as noteworthy temporal trends over the past decade. Specifically, between 2018 and 2021, dementia garnered an increasing amount of public attention on social media. In particular, exponential growth in dementia-related posts has been occurring since 2018. Several explanations for this temporal shift are possible. First, the drastic rise in public attention being paid to dementia could be attributed partly to the celebrity effect. For instance, Sir Charles Kuen Kao, the father of fiber optics and recipient of the Nobel Prize in Physics, died in September 2018 following a decade-long battle with Alzheimer disease. Kao and his family established the Charles K. Kao Foundation for Alzheimer's Disease in 2010 to raise public awareness about Alzheimer disease and dementia in Chinese society. The Kao family's public disclosure received extensive coverage in Chinese media, which may have led to the growing trend since 2018. Indeed, the results from the time series analysis using the time of Sir Kao's death as the time of event indicated significant changes in monthly dementia post counts (see Multimedia Appendix 1, "Interrupted Time Series," for detailed results). Similarly, a reality show "Forget Me Not Café," starring 5 older people with dementia, premiered in China in 2019 to raise public awareness about dementia. The show received around 6.7 million views between its debut and February 2021 [47]. Second, a succession of dementia-related initiatives in China also may have contributed to increased discussion of dementia on social media [10,48]. For instance, Alzheimer's Disease China (ADC) was founded in 2002 to serve as the Alzheimer's Disease International's Chinese affiliate. In 2011, China's 12th Five-Year Development Plan for Population Aging recognized the need for early dementia intervention. In 2014, the first Memory Clinic Guide was published to aid in the timely diagnosis and treatment of dementia. In 2018, about 100 memory clinics were in operation. Furthermore, the National Health Commission of the People's Republic of China launched the *Alzheimer's Disease Prevention and Treatment Guide* in 2019 [49]. Collectively, these national initiatives may have contributed greatly to an upsurge in dementia conversations on social media.

Our results indicate that the derogatory term for dementia in China, "老年痴呆," which roughly translates to "old and stupid," has remained the most frequently used term for dementia on social media over the past decade. Although alternative terms for dementia have been introduced and advocated in China [50], this derogatory term continues to dominate public discourse. Considering that existing research suggests that derogatory local terms for dementia may result in strong stigmatization and, thus, discourage people from seeking help, multiple East Asian nations (eg, Japan, Hong Kong, and Singapore) have launched campaigns to replace their negative local terms for dementia with more neutral biomedical ones [51], such as "cognitive disorder" ("認知障礙症") and "brain degeneration syndrome" ("腦退化症") in Hong Kong and "memory disorder" ("失智症") in Singapore and Taiwan. A national effort to rename dementia terms is essential in helping to change public views and attitudes toward dementia in Mainland China. Nevertheless, it is worth noting that advocacy and policy have been gaining momentum on Chinese social media since 2019. Collectively, these results indicate that although dementia has been linked with stigmatizing local terms, the future looks bright, as advocacy- and policy-related discussions gain momentum.

Another noteworthy finding is that raising awareness (ie, posts promoting greater understanding and positive attitudes toward dementia) has remained the least discussed dementia-related subject on social media. For example, a Weibo user published a post that read, "To appeal to society, many elderly people in rural areas are not simply old and confused; they are just sick," highlighting the need to eradicate prejudice toward older adults with dementia. This is concerning, given the overwhelming evidence supporting the importance of raising awareness to promote early detection and treatment of dementia [10,52]. These findings underscore the urgent need for public health campaigns on dementia in China. Improved awareness and knowledge of dementia may help reduce stigmatization and discrimination, as well as enhance social support for affected individuals and their families [53,54], thereby helping to develop dementia-friendly communities in which families affected by dementia are supported and included [55]. Given the surge in social media use, leveraging social media outlets offers a cost-effective way to raise public awareness of dementia (eg, sharing dementia experiences, stories, and knowledge via public messaging and general communication). In some countries (eg, Australia), media campaigns have been critical in gaining government recognition of dementia as a national health priority [53].

The consistent results from 2 typical text-mining methods (ie, topic modeling and semantic network analysis) indicate our findings' robustness. Our findings suggest that symptoms, treatment, and prevention, which often indicate deficits and incapacities, are the most discussed aspects of dementia overall. The dominance of such medical discourse may exacerbate stigmatization and fear toward those with dementia [56,57]. Semantic analysis further revealed that various Weibo user groups focus on distinct aspects of dementia. Specifically, government users focused primarily on prevention, social support, and symptom management, while prevention, treatment, and symptoms dominated scientists/experts' discourse. Media coverage, personal experiences, and social support dominated journalists/news media's online discourse, while personal experience and social support aspects were relatively more prominent in the general public's discourse. Furthermore, dementia frame adoption differed significantly across the 4 user groups. Government users and journalists/news media users had a greater degree of agreement on dementia discourse, with a particular emphasis on media coverage, prevention, and social support. However, the general public and scientists/experts mostly emphasized treatment, prevention, and social support.

Regarding sentiment toward dementia, our findings suggest that the general public is more prone than other user groups (ie, the government, journalists/news media, and scientists/experts) toward harboring negative sentiments toward dementia. Furthermore, compared with posts on personal experiences, only microblogging posts with a policy and advocacy focus indicated positive sentiments. All the remaining dementia frames/themes were associated with negative sentiments. Furthermore, an interesting temporal trend was discovered. Specifically, our results reveal that compared with the baseline in 2010, sentiments toward dementia on Chinese social media were more positive from 2011 to 2017. Collectively, sentiment analyses indicate that even though sentiment toward dementia has become more positive over the years, it remains negative overall, particularly among the general public. Owing to the availability of advanced analytical tools, these findings add an emotional dimension to our understanding of public opinion on dementia in China.

Furthermore, this study's results further reveal that a variety of user groups (eg, the general public, health care providers, and policymakers) utilize Weibo to receive and share information related to dementia. This finding corresponds with earlier research indicating that social media platforms are the primary information-seeking and information-sharing medium for dementia discourse [19,42]. Consequently, Weibo could serve as a medium for dementia awareness-raising and educational initiatives.

## Implications

The study's findings carry significant implications for research, policy, and practice. First, considering diverse user groups' active engagement and the abundance of microblogging content concerning dementia, Weibo could be leveraged as a primary channel through which to disseminate dementia-related educational programs to promote awareness and positive attitudes toward dementia. Future research needs to evaluate whether Weibo could be an effective modality for engaging the general public in dementia education and other health promotion activities related to dementia. Second, despite national policy and awareness-raising campaigns, the most widely used Chinese term for dementia is still one that carries considerable negative connotations. Thus, a renaming campaign is needed urgently in Mainland China. Third, the results reveal widespread unfavorable attitudes toward dementia, particularly among the general public, highlighting the critical need for public health campaigns with multiprong approaches (eg, academic institutions collaborating with community organizations and health care delivery organizations, including hospitals and clinics, to produce more dementia-friendly films, documentaries, reality shows, and educational brochures). Furthermore, computational social science is gaining prominence. Advanced information processing analytical tools (eg, machine learning) are being used to analyze vast amounts of data to acquire a better understanding of social and public health issues. Future studies need to deepen similar public health–computational social science collaborations to advance our knowledge of a variety of social challenges. Moreover, our findings reveal that dementia is connected constantly with negative sentiments, with symptoms and prevention dominating public discourse. Developing interventions to combat such an overtly unfavorable view of dementia requires policy and public health attention. Finally, although keywords used in this study are in line with common practices in the field, new terms or words that refer to dementia may emerge on social media as online discourses may evolve over time. Future researchers can build on our keyword list and actively monitor and add new keywords to develop an enhanced and dynamic understanding of dementia (eg, leveraging "keywords-as-frames" as the discourse analysis tool in future studies).

## Limitations and Strengths

Several limitations of the study warrant discussion. First, considering that the data used in this study were obtained only from Weibo users, the study's findings may not be generalizable to the entire Mainland China's population (eg, those without internet access and non-Weibo users). Certain populations' voices (eg, those with dementia and their caregivers) may not be represented in this study, owing to possible difficulties with social media navigation. The lived experiences of people with dementia and their family caregivers require further investigation to refute prevailing negative and biomedical discourses about dementia [57]. Moreover, it is possible that some extreme views on dementia and negative sentiments may not be available on these social media platforms; thus, such views may not have been captured in this study. Nevertheless, considering that Weibo is China's most popular social media site, the study's findings reflect population-level perceptions of dementia in China. Future studies using large-scale data from other sources are needed to corroborate our results. Furthermore, this study examined the public perception of dementia only in Mainland China; thus, further research into cross-cultural comparisons of social media representations of dementia in other countries/cultures is warranted.

Despite these limitations, this study overcomes other limitations inherent in survey research on dementia, in which realizing a national understanding of public discourse and opinion concerning dementia remains challenging. The findings built on existing research by adding a macrolevel understanding of the Chinese public's discourse and sentiment toward dementia, which is critical for developing national education and policy initiatives to create a dementia-friendly society. Furthermore, our findings respond to the global call to expand dementia research in low- and middle-income countries [53].

## Conclusion

Dementia-related discussions have expanded rapidly on social media in China during the past decade, most notably since 2018. However, a derogatory local Chinese term for dementia remains in common use; thus, a national renaming effort is imperative to reform the public discourse on dementia. Discussions related to dementia advocacy and policy are gaining momentum in China, but promoting awareness of dementia has remained the least discussed topic over time. The sentiment toward dementia is negative overall in social media, particularly among the general public, but social media platforms are part of a viable medium for providing dementia education and awareness-raising interventions in Mainland China.

## Acknowledgments

DK, JZ, and AC contributed to the study's design. AC collected Weibo data. DK, JZ, and AC contributed to data analysis. DK and AC contributed to the design and drawing of figures. All authors contributed to the writing and revising of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection and preprocessing.
[DOCX File , 263 KB - jmir_v24i9e39805_app1.docx ]

## References

1. Prince M, Wimo A, Guerchet M, Ali G, Wu Y, Prina M. World Alzheimer Report 2015-The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost, and Trends. 2015. URL: https://www.alzint.org/u/WorldAlzheimerReport2015.pdf [accessed 2022-08-17]
2. Wu Y, Ali G, Guerchet M, Prina AM, Chan KY, Prince M, et al. Prevalence of dementia in Mainland China, Hong Kong and Taiwan: an updated systematic review and meta-analysis. Int J Epidemiol 2018 Jun 01;47(3):709-719 [FREE Full text] [doi: 10.1093/ije/dyy007] [Medline: 29444280]
3. Xu J, Wang J, Wimo A, Fratiglioni L, Qiu C. The economic burden of dementia in China, 1990-2030: implications for health policy. Bull World Health Organ 2017 Jan 01;95(1):18-26 [FREE Full text] [doi: 10.2471/BLT.15.167726] [Medline: 28053361]
4. Chen R, Hu Z, Chen R, Ma Y, Zhang D, Wilson K. Determinants for undetected dementia and late-life depression. Br J Psychiatry 2013 Sep;203(3):203-208. [doi: 10.1192/bjp.bp.112.119354] [Medline: 23888000]
5. Zhao M, Lv X, Tuerxun M, He J, Luo B, Chen W, Clinical Pathway for Alzheimer’s Disease in China (CPAD) investigators. Delayed help seeking behavior in dementia care: preliminary findings from the Clinical Pathway for Alzheimer's Disease in China (CPAD) study. Int Psychogeriatr 2016 Feb;28(2):211-219. [doi: 10.1017/S1041610215000940] [Medline: 26138923]
6. Jia L, Quan M, Fu Y, Zhao T, Li Y, Wei C, et al. Dementia in China: epidemiology, clinical management, and research advances. Lancet Neurol 2020 Jan;19(1):81-92. [doi: 10.1016/S1474-4422(19)30290-X] [Medline: 31494009]
7. Jia J, Zuo X, Jia X, Chu C, Wu L, Zhou A, China Cognition and Aging Study (China COAST) Group. Diagnosis and treatment of dementia in neurology outpatient departments of general hospitals in China. Alzheimers Dement 2016 Apr;12(4):446-453. [doi: 10.1016/j.jalz.2015.06.1892] [Medline: 26256457]
8. de Vugt ME, Verhey FRJ. The impact of early dementia diagnosis and intervention on informal caregivers. Prog Neurobiol 2013 Nov;110:54-62. [doi: 10.1016/j.pneurobio.2013.04.005] [Medline: 23689068]
9. Leifer B. Early diagnosis of Alzheimer's disease: clinical and economic benefits. J Am Geriatr Soc 2003 May;51(5 Suppl Dementia):S281-S288. [doi: 10.1046/j.1532-5415.5153.x] [Medline: 12801384]
10. Chen Z, Yang X, Song Y, Song B, Zhang Y, Liu J, et al. Challenges of dementia care in China. Geriatrics (Basel) 2017 Jan 18;2(1):7 [FREE Full text] [doi: 10.3390/geriatrics2010007] [Medline: 31011017]
11. Li X, Fang W, Su N, Liu Y, Xiao S, Xiao Z. Survey in Shanghai communities: the public awareness of and attitude towards dementia. Psychogeriatrics 2011 Jun;11(2):83-89 [FREE Full text] [doi: 10.1111/j.1479-8301.2010.00349.x] [Medline: 21707855]
12. Liu D, Cheng G, An L, Gan X, Wu Y, Zhang B, et al. Public knowledge about dementia in China: a national WeChat-based survey. Int J Environ Res Public Health 2019 Oct 31;16(21):4231 [FREE Full text] [doi: 10.3390/ijerph16214231] [Medline: 31683665]
13. Zhang H, Loi S, Zhou S, Zhao M, Lv X, Wang J, et al. Dementia literacy among community-dwelling older adults in urban China: a cross-sectional study. Front Public Health 2017;5:124 [FREE Full text] [doi: 10.3389/fpubh.2017.00124] [Medline: 28638820]
14. Hsiao H, Liu Z, Xu L, Huang Y, Chi I. Knowledge, attitudes, and clinical practices for patients with dementia among mental health providers in China: city and town differences. Gerontol Geriatr Educ 2016;37(4):342-358. [doi: 10.1080/02701960.2014.990152] [Medline: 25625718]
15. Zhao W, Jones C, Wu M, Moyle W. Healthcare professionals' dementia knowledge and attitudes towards dementia care and family carers' perceptions of dementia care in China: an integrative review. J Clin Nurs 2022 Jul 31;31(13-14):1753-1775. [doi: 10.1111/jocn.15451] [Medline: 32786146]
16. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]
17. China Internet Network Information Center (CNNIC). The 46th Survey Report. URL: http://www1.cnnic.cn/IDR/ReportDownloads/201302/P020130221391269963814.pdf [accessed 2022-08-17]
18. Zhang X, Wen D, Liang J, Lei J. How the public uses social media WeChat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
19. Leung AYM, Molassiotis A, Zhang J, Deng R, Liu M, Van IK, et al. Dementia literacy in the Greater Bay Area, China: identifying the at-risk population and the preferred types of mass media for receiving dementia information. Int J Environ Res Public Health 2020 Apr 07;17(7):2511 [FREE Full text] [doi: 10.3390/ijerph17072511] [Medline: 32272551]
20. Mhasawade V, Zhao Y, Chunara R. Machine learning and algorithmic fairness in public and population health. Nat Mach Intell 2021 Jul 29;3(8):659-666. [doi: 10.1038/s42256-021-00373-4]
21. Dredze M. How social media will change public health. IEEE Intell Syst 2012 Jul;27(4):81-84 [FREE Full text] [doi: 10.1109/MIS.2012.76]
22. Guan W, Gao H, Yang M, Li Y, Ma H, Qian W, et al. Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events. Phys A: Stat Mech Its Appl 2014 Feb;395:340-351. [doi: 10.1016/j.physa.2013.09.059]
23. Lam WS. Watch the Media Controversy of “Rectification” of Special Geriatrics. URL: http://gbcode.rthk.hk/TuniS/app3.rthk.hk/mediadigest/content.php?aid=1198 [accessed 2022-08-17]

24. Chen A, Zhang X. Changing social representations and agenda interactions of gene editing after crises: a network agenda-setting study on Chinese social media. Soc Sci Comput Rev 2021 Mar 15:089443932199806. [doi: 10.1177/0894439321998066]
25. Shahin S. User-generated nationalism: interactions with religion, race, and partisanship in everyday talk online. Info Commun Soc 2020 Jun 22;24(13):1854-1869. [doi: 10.1080/1369118x.2020.1748088]
26. Chen K, Tomblin D. Using data from Reddit, public deliberation, and surveys to measure public opinion about autonomous vehicles. Public Opin Q 2021;85(1):322. [doi: 10.1093/poq/nfab021]
27. Li P, Cho H, Qin Y, Chen A. #MeToo as a connective movement: examining the frames adopted in the anti-sexual harassment movement in China. Soc Sci Comput Rev 2020 Sep 15;39(5):1030-1049. [doi: 10.1177/0894439320956790]
28. BARBERÁ P, CASAS A, NAGLER J, EGAN PJ, BONNEAU R, JOST JT, et al. Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. Am Polit Sci Rev 2019 Jul 12;113(4):883-901. [doi: 10.1017/s0003055419000352]
29. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. Adv Neural Inf Process Syst 2002;3:993.
30. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]
31. Arun R, Suresh V, Madhavan C, Murty M. On finding the natural number of topics with latent Dirichlet allocation: some observations. In: Zaki MJ, Yu JX, Ravindran B, Pudi V, editors. Advances in Knowledge Discovery and Data Mining. PAKDD 2010. Lecture Notes in Computer Science. Volume 6118. Berlin, Heidelberg: Springer; 2010.
32. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. Neurocomputing 2009 Mar;72(7-9):1775-1781. [doi: 10.1016/j.neucom.2008.06.011]
33. Griffiths TL, Steyvers M. Finding scientific topics. Proc Natl Acad Sci U S A 2004 Apr 06;101 Suppl 1:5228-5235 [FREE Full text] [doi: 10.1073/pnas.0307752101] [Medline: 14872004]
34. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Doc Numér 2014 Apr 30;17(1):61-84. [doi: 10.3166/dn.17.1.61-84]
35. Nelson LK. Computational grounded theory: a methodological framework. Sociol Methods Res 2017 Nov 21;49(1):3-42. [doi: 10.1177/0049124117729703]
36. Luo C, Chen A, Cui B, Liao W. Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: semantic network analysis of two social media platforms in the United States and China. Telemat Inform 2021 Dec;65:101712 [FREE Full text] [doi: 10.1016/j.tele.2021.101712] [Medline: 34887618]
37. Xi Y, Chen A, Zhang W. The expression of cultural identities in Hong Kong's Anti-Extradition Law Amendment Bill movement: a semantic network analysis of tweets. Soc Sci Comput Rev 2021 May 03:089443932110122. [doi: 10.1177/08944393211012267]
38. Ruiz JB, Barnett GA. Exploring the presentation of HPV information online: a semantic network analysis of websites. Vaccine 2015 Jun 26;33(29):3354-3359. [doi: 10.1016/j.vaccine.2015.05.017] [Medline: 26036946]
39. Zhang L, Xu L, Zhang W. Social media as amplification station: factors that influence the speed of online public response to health emergencies. Asian J Commun 2017 Feb 20;27(3):322-338. [doi: 10.1080/01292986.2017.1290124]
40. Liu J, Seneff S. Review sentiment scoring via a parse-and-paraphrase paradigm. 2009 Presented at: EMNLP '09: 2009 Conference on Empirical Methods in Natural Language Processing; August 2009; Singapore. [doi: 10.3115/1699510.1699532]
41. Subrahmanian V, Reforgiato D. AVA: adjective-verb-adverb combinations for sentiment analysis. IEEE Intell Syst 2008 Jul;23(4):43-50. [doi: 10.1109/mis.2008.57]
42. Wang Z, Zou N, Xie B, Luo Z, He D, Hilsabeck R, et al. Characterizing dementia caregivers' information exchange on social media: exploring an expert-machine co-development process. In: Toeppe K, Yan H, Chu SKW, editors. Diversity, Divergence, Dialogue. iConference 2021. Lecture Notes in Computer Science. Volume 12645. Cham: Springer; 2021.
43. Sm-Rahman A, Lo CH, Jahan Y. Dementia in media coverage: a comparative analysis of two online newspapers across time. Int J Environ Res Public Health 2021 Oct 08;18(19):10539 [FREE Full text] [doi: 10.3390/ijerph181910539] [Medline: 34639840]
44. Petersen N, Schicktanz S. The experts' advice: prevention and responsibility in German media and scientific discourses on dementia. Qual Health Res 2021 Sep;31(11):2005-2018 [FREE Full text] [doi: 10.1177/10497323211014844] [Medline: 34109877]
45. Yoon S. What can we learn about mental health needs from tweets mentioning dementia on World Alzheimer's Day? J Am Psychiatr Nurses Assoc 2016 Nov;22(6):498-503 [FREE Full text] [doi: 10.1177/1078390316663690] [Medline: 27803262]
46. Kirkman AM. Dementia in the news: the media coverage of Alzheimer's disease. Australas J Ageing 2006 Jun;25(2):74-79. [doi: 10.1111/j.1741-6612.2006.00153.x]
47. Xu W. The stigma of dementia and the media: an analysis of reality shows about older people with dementia running a pop-up restaurant. J Aging Stud 2021 Dec;59:100967 [FREE Full text] [doi: 10.1016/j.jaging.2021.100967] [Medline: 34794712]
48. Lou VWQ, Chi I. Is China ready for dementia? J Alzheimer's Neurodegener Dis 2020 Jun 30;6(2):1-2. [doi: 10.24966/and-9608/100041]

49. Chan KY, Wang W, Wu JJ, Liu L, Theodoratou E, Car J, et al. Epidemiology of Alzheimer's disease and other forms of dementia in China, 1990–2010: a systematic review and analysis. Lancet 2013 Jun 08;381(9882):2016-2023 [FREE Full text] [doi: 10.1016/S0140-6736(13)60221-4] [Medline: 23746902]
50. Dai B, Mao Z, Wu B, Mei YJ, Levkoff S, Wang H. Family caregiver's perception of Alzheimer's disease and caregiving in Chinese culture. Soc Work Public Health 2015;30(2):185-196. [doi: 10.1080/19371918.2014.969858] [Medline: 25602761]
51. Chiu HFK, Sato M, Kua EH, Lee M, Yu X, Ouyang W, et al. Renaming dementia--an East Asian perspective. Int Psychogeriatr 2014 Jun;26(6):885-887. [doi: 10.1017/S1041610214000453] [Medline: 24697867]
52. Low L, Anstey KJ. Dementia literacy: recognition and beliefs on dementia of the Australian public. Alzheimers Dement 2009 Jan;5(1):43-49. [doi: 10.1016/j.jalz.2008.03.011] [Medline: 19118808]
53. Prince M, Acosta D, Albanese E, Arizaga R, Ferri CP, Guerra M, et al. Ageing and dementia in low and middle income countries: using research to engage with public and policy makers. Int Rev Psychiatry 2008 Aug;20(4):332-343 [FREE Full text] [doi: 10.1080/09540260802094712] [Medline: 18925482]
54. Wang S, Cheung DSK, Leung AYM. Overview of dementia care under the three-tier long-term care system of China. Public Health Nurs 2019 Mar;36(2):199-206. [doi: 10.1111/phn.12573] [Medline: 30549090]
55. Rahman S, Swaffer K. Assets-based approaches and dementia-friendly communities. Dementia (London) 2018 Feb 04;17(2):131-137. [doi: 10.1177/1471301217751533] [Medline: 29299934]
56. Zeilig H. Dementia as a cultural metaphor. Gerontologist 2014 Apr;54(2):258-267. [doi: 10.1093/geront/gns203] [Medline: 23408265]
57. Swaffer K. Dementia: stigma, language, and dementia-friendly. Dementia (London) 2014 Nov 17;13(6):709-716. [doi: 10.1177/1471301214548143] [Medline: 25326220]

## Abbreviations

**LDA:** latent Dirichlet allocation

*Edited by A Mavragani; submitted 24.05.22; peer-reviewed by J Xu, SF Tsao, D Helic; comments to author 16.06.22; revised version received 15.07.22; accepted 23.07.22; published 02.09.22.*

*Please cite as:*
*Kong D, Chen A, Zhang J, Xiang X, Lou WQV, Kwok T, Wu B*
*Public Discourse and Sentiment Toward Dementia on Chinese Social Media: Machine Learning Analysis of Weibo Posts*
*J Med Internet Res 2022;24(9):e39805*
*URL: https://www.jmir.org/2022/9/e39805*
*doi:10.2196/39805*
*PMID:36053565*

Original Paper

# Quality and Audience Engagement of Takotsubo Syndrome–Related Videos on TikTok: Content Analysis

Jing Liang[1], MD; Linlin Wang[2], MD; Shijie Song[3], PhD; Man Dong[4], MD; Yidan Xu[1], MD; Xinyu Zuo[1], MD; Jingyi Zhang[1], MD; Akil Adrian Sherif[5], MD; Jafree Ehsan[6], MD; Jianjun Ma[1*], MD; Pengyang Li[7*], MD

[1]Xinxiang Medical University, Henan, China
[2]Aerospace Central Hospital, Beijing, China
[3]Business School, Hohai University, Nanjing, China
[4]The Third Affiliated Hospital of Xinxiang Medical University, Henan, China
[5]Divison of Cardiology, Saint Vincent Hospital, Worcester, MA, United States
[6]Department of Medicine, Virginia Commonwealth University, Richmond, VA, United States
[7]Division of Cardiology, Pauley Heart Center, Virginia Commonwealth University, Richmond, VA, United States
[*]these authors contributed equally

**Corresponding Author:**
Pengyang Li, MD
Division of Cardiology, Pauley Heart Center
Virginia Commonwealth University
1200 E Marshall St
Richmond, VA, 23219
United States
Phone: 1 626 420 5811
Email: leelpy0109@gmail.com

## Abstract

**Background:** The incidence of Takotsubo syndrome (TTS), also known as the broken heart syndrome or stress cardiomyopathy, is increasing worldwide. The understanding of its prognosis has been progressively evolving and currently appears to be poorer than previously thought, which has attracted the attention of researchers. An attempt to recognize the awareness of this condition among the general population drove us to analyze the dissemination of this topic on TikTok, a popular short-video–based social media platform. We found a considerable number of videos on TTS on TikTok; however, the quality of the presented information remains unknown.

**Objective:** The aim of this study was to analyze the quality and audience engagement of TTS-related videos on TikTok.

**Methods:** Videos on the TikTok platform were explored on August 2, 2021 to identify those related to TTS by using 6 Chinese keywords. A total of 2549 videos were found, of which 80 met our inclusion criteria and were evaluated for their characteristics, content, quality, and reliability. The quality and reliability were rated using the DISCERN instrument and the Journal of the American Medical Association (JAMA) criteria by 2 reviewers independently, and a score was assigned. Descriptive statistics were generated, and the Kruskal-Wallis test was used for statistical analysis. Multiple linear regression was performed to evaluate the association between audience engagement and other factors such as video content, video quality, and author types.

**Results:** The scores assigned to the selected video content were low with regard to the diagnosis (0.66/2) and management (0.34/2) of TTS. The evaluated videos were found to have an average score of 36.93 out of 80 on the DISCERN instrument and 1.51 out of 4 per the JAMA criteria. None of the evaluated videos met all the JAMA criteria. The quality of the relayed information varied by source (All $P<.05$). TTS-related videos made by health care professionals accounted for 28% (22/80) of all the evaluated videos and had the highest DISCERN scores with an average of 40.59 out of 80. Multiple linear regression analysis showed that author types that identified as health professionals (exponentiated regression coefficient 17.48, 95% CI 2.29-133.52; $P$=.006) and individual science communicators (exponentiated regression coefficient 13.38, 95% CI 1.83-97.88; $P$=.01) were significant and independent determinants of audience engagement (in terms of the number of likes). Other author types of videos, video content, and DISCERN document scores were not associated with higher likes.

**Conclusions:** We found that the quality of videos regarding TTS for patient education on TikTok is poor. Patients should be cautious about health-related information on TikTok. The formulation of a measure for video quality review is necessary, especially when the purpose of the published content is to educate and increase awareness on a health-related topic.

***(J Med Internet Res 2022;24(9):e39360)*** doi:10.2196/39360

**KEYWORDS**

TikTok; short video apps; information quality; Takotsubo syndrome; patient education; social media; audience engagement

## *Introduction*

Takotsubo syndrome (TTS), also known as the broken heart syndrome or stress cardiomyopathy, is characterized by transient ventricular dysfunction with typical wall motion abnormalities [1,2]. The number of patients diagnosed with TTS has been gradually increasing [3], with hospitalizations for TTS increasing from 5.7 per 100,000 person-years in 2007 to 17.4 per 100,000 person-years in 2012 ($P<.001$) [4]. The clinical manifestation of TTS is similar to that of acute coronary syndrome (ACS) and frequently presents with chest discomfort or dyspnea, ST segment deviation on electrocardiogram, and cardiac biomarker abnormalities [2,5,6]. Given the overlap with the clinical presentation of ACS, TTS can be easily misdiagnosed. Previous reports [7,8] have shown that about 0.7%-2% of all patients with possible ACS were eventually found to have had TTS. Ongoing research has shown that the prognosis of TTS is not as benign as previously thought, with in-hospital mortality of 4.1% and long-term mortality of 24.7% [6,9].

The internet is a useful platform for effectively communicating new information, and many new technological applications have taken advantage of this to serve as a medium for patient education on health-related topics. Seeking web-based health information has become increasingly popular; many people rely on web-based resources to obtain health information and aid their medical decision-making [10]. Research has shown that health outcomes can be positively impacted by appropriately and effectively utilizing social media platforms [11]. One of the many social media applications that has been used to reach a large audience is TikTok [12]. However, the most important limitation of such platforms is the unreliable quality of the information presented. Anyone can present information on social media platforms, most of which lack formal moderation for authenticity and reliability of the presented material [13]. Recent systematic reviews have suggested that the quality of web-based health information is problematic and perhaps made worse when considering information disseminated on social media platforms [14].

Along with the quality of information, audience engagement is another key component of effective web-based health communication. Audience engagement in health-related topics has been studied in traditional social media platforms such as YouTube and Facebook [15-19]. The influence of many factors such as video content, quality, and information sources on audience engagement has been studied previously [15,16,18-20]. Several studies have shown that video content is associated with audience engagement [15,18,20]. Szmuda et al [15] studied the association between video content and audience engagement in COVID-19 videos on YouTube, and they found that videos showing the causes, management strategies, diagrams, and structure anatomies were associated with a higher “like” ratio. Another study [20] on audience engagement and COVID-19 short videos on TikTok found that content type (news, codebook, etc) influenced the level of audience engagement.

Previous studies have yielded inconsistent conclusions on the impact of video quality on audience engagement. A study on videos of stroke on YouTube by Szmuda et al [16] showed that there was no strong correlation between the DISCERN score (an indicator of video quality) and audience engagement. Huang et al [19] found that nephrolithiasis-related videos on YouTube with inaccurate statements were associated with higher audience engagement (viewer-generated comments, thumbs-up and thumbs-down ratings) than videos without inaccuracies. The source of information was also found to affect audience engagement. Szmuda et al [16] found that higher engagement was noted in stroke-related videos that were uploaded by an educational channel on YouTube. Recently, studies [21,22] have shown that the emerging short-video social apps can satisfy people’s intrinsic motivations and elicit user engagement when disseminating health information.

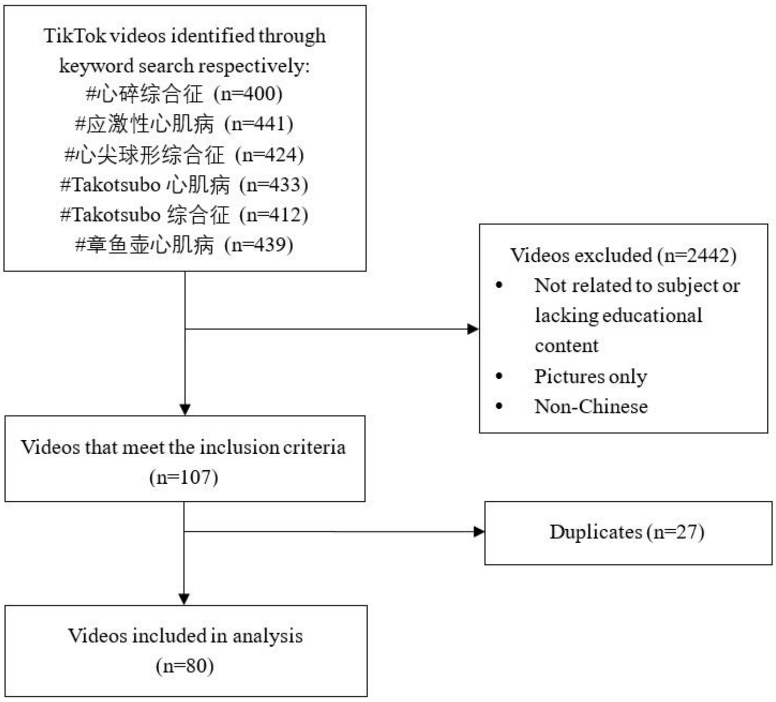
TikTok (DouYin in Chinese) is a short-video social app with a sizable userbase wherein individual users create and publicly post short videos on various subjects. Initially, when the platform first became popular, the video length was limited to short 60-second clips. However, with growing demand, the length was extended to allow up to 5-minute-long videoclips [23]. TikTok has attained significant popularity since its launch in September 2016 and has since raked up more than 500 million active users and 1 billion downloads [24,25]. Given its extensive reach and better audience engagement than other traditional social media platforms [23], TikTok has the potential to be a great source of health information dissemination and become increasingly popular among general health consumers as an emerging health information source. Several studies evaluating patient education videos regarding COVID-19 [26], skin-related diseases [27], chronic obstructive pulmonary disease [28], and aesthetic surgery procedures [29] posted on TikTok have been published. These studies have revealed that the overall quality of such videos presenting health information on TikTok is low, and some even present overtly false information [27,29]. Very few studies [15,16,18,20] have focused on audience engagement with health care information on TikTok as compared to that with health care information on other social media platforms. We found a considerable number of videos regarding TTS on TikTok; however, the quality and content of the presented information and whether these factors may affect audience engagement remain largely unknown. Therefore, this study aims to evaluate the content and quality of videos related to TTS on TikTok and assess the qualitative metrics that drive audience engagement (in terms of the number of likes) with a video.

## Methods

### Search Strategy

The search was conducted on August 2, 2021, in China. Six specific hashtags（“#应激性心肌病” “#Takotsubo综合征” “#Takotsubo心肌病” “#心碎综合征” “#章鱼壶心肌病” “#心尖球形综合征”）that refer to TTS in Chinese were used to retrieve TikTok videos related to TTS. TikTok provides 3 ways to filter videos, that is, overall ranking, most recent, and most liked. Considering that most users use the default sorting option, “overall ranking,” we performed the search in TikTok under the discover mode using the “overall ranking” sort option. All the resultant videos for each keyword were retrieved and screened. The initial search returned 2549 videos (441, 400, 433, 412, 439, and 424, respectively), of which 80 videos met our criteria for analysis after screening. The exclusion criteria were as follows: (1) videos not related to TTS or lacking educational information, (2) duplicate videos, (3) videos not in Chinese, and (4) videos that were not original. Figure 1 illustrates the selection process implemented in our study.

**Figure 1.** Selection process implemented in this study.



### Data Extraction

Baseline characteristics were extracted directly from each video and the video author’s public TikTok profile on the same day of the data search. With respect to the video authors, we collected their account ID, inauguration unit, number of both followers and those they follow, overall number of posted videos, and likes obtained. Additionally, we further ascertained whether the public TikTok profile had photos, live broadcast information, and contact details. For each individual video, we also collected the uniform resource locator, the date posted, the number of likes, comments, reposts, and duration. Based on the abovementioned information and historical videos on their profile page, the TikTok authors were classified into 6 categories: (1) individual science communicators, (2) news agencies, (3) for-profit organizations, (4) health professionals, (5) nonprofit organizations, and, (6) general users. Furthermore, the content of each video was assessed for the following characteristics: use of characters, background music, emoji, flash, and video subtitles.

### Coding Schema

We used the DISCERN instrument and Journal of the American Medical Association (JAMA) criteria to evaluate the quality of the selected TikTok videos. These instruments are commonly used standards for health information evaluation [30,31]. The DISCERN instrument (quality criteria for consumer health information) consists of 16 questions, with each question scored from 1 to 5 points. Questions are divided into 3 parts: reliability of the publication (questions 1 to 8), quality of information about treatment options (questions 9 to 15), and the overall score of the publication (question 16) [32]. The JAMA benchmarks were applied to evaluate health-related information’s reliability, plausibility, and usefulness in the internet [33]. The JAMA criteria consist of 4 main sections, which are scored from 0 to 4. Six questions obtained from reports of Goobie et al [34] were used to measure the quality of the video content. The 6 questions assessed the definition, signs or symptoms, risk factors, evaluation, management, and outcomes of the disease discussed in the videos. Each aspect was scored across 3 degrees from 0 to 2: not addressed, partially addressed, and sufficiently addressed.

### Video Coding

All video contents evaluated using the DISCERN instrument, JAMA benchmarks, and the 6 questions from the studies of Goobie et al [34] were independently scored and coded by 2 study authors (JL and YX). Prior to coding, a training exercise was conducted, during which 20 videos each were scored and coded independently by the 2 coding authors, and any resulting discrepancies were addressed and resolved to ensure homogeneity in coding. An average of 2 points was assigned by rounding to the nearest unit. Microsoft Excel (2019) was used to extract, code the basic information of each video, and process statistical data.

### Statistical Analysis

Descriptive statistics for continuous variables, including the mean, median, range, and standard deviation, were calculated. SPSS version 22.0 (IBM Corp) was used to perform data analysis. An intraclass correlation coefficient was used to assess the interrater agreement. The average agreement by intraclass correlation coefficient was 0.87 for content rating, 0.92 for DISCERN scales rating, and 0.88 for the JAMA benchmark rating. The average intraclass correlation coefficients for interrater agreements all exceeded the recommended value of 0.75, indicating that the ratings had good reliability [35]. The Kruskal-Wallis test was performed to identify differences between the extracted variables. A *P* value <.05 was deemed statistically significant. In our paper, we use the number of likes as a proxy measure of audience engagement. Multiple linear regression was performed to evaluate the association between audience engagement and other factors such as video quality and author types. The regression analysis was conducted on R (R version 4.2.1; 2022-06-23 ucrt), with a log transformation of the outcome variable audience engagement. This study did not involve human subjects, and hence, the study was not reviewed by the institutional review board.

## Results

### Video Characteristics

The average length of the evaluated videos was 59 seconds, with a maximum duration of 700 seconds and a minimum duration of 7 seconds. On average, a video received 5543 likes and 635 reposts. The majority of the videos (49/80, 61%) had the presence of people, 90% (72/80) had subtitles, and 70% (56/80) of the videos had background music. Approximately 19% (15/80) and 23% (18/80) of the videos had emojis and animations, respectively (Table 1).

With regard to content creation, users classified as health professionals posted the most videos (22/80, 28%), followed by general users (20/80, 25%), individual science communicators (18/80, 23%), news agencies (11/80, 14%), and for-profit organizations (6/80, 8%). Nonprofit organizations posted the fewest videos (3/80, 4%) (Multimedia Appendix 1). The average number of likes received per video, categorized by video author types, from high to low were as follows: news agencies, health professionals, for-profit organizations, individual science communicators, general users, and nonprofit organizations (Table 2).

**Table 1.** Basic characteristics of the Takotsubo syndrome–related videos on TikTok.

| | Mean | Median |
|---|---|---|
| **Author's account information** | | |
| Likes | 3,451,837 | 188,000 |
| Following | 166 | 48 |
| Followers | 622,697 | 41,000 |
| Videos | 743 | 141 |
| **Characteristics of Takotsubo syndrome–related videos** | | |
| Duration (min) | 59 | 41 |
| Likes | 5543 | 108 |
| Comments | 518 | 8 |
| Reposts | 635 | 8 |
| **Information quality** | | |
| Six questions | 5.40 | N/A[a] |
| DISCERN score | 36.93 | N/A |
| Journal of the American Medical Association criteria | 1.51 | N/A |

[a] N/A: not applicable.

**Table 2.** Characteristics of Takotsubo syndrome–related videos on TikTok categorized by source.

| | Health professionals (n=22) | Individual science communicators (n=18) | General users (n=20) | News agencies (n=11) | For-profit organizations (n=6) | Nonprofit organizations (n=3) |
|---|---|---|---|---|---|---|
| **Author's account information (mean)** | | | | | | |
| Likes | 1,764,734 | 2,029,657 | 115,554 | 17,184,623 | 4,066,700 | 61,000 |
| Following | 168 | 64 | 343 | 104 | 114 | 19 |
| Followers | 1,513,977 | 192,388 | 26,012 | 1,022,507 | 288,864 | 12,831 |
| Videos | 319 | 101 | 97 | 3969 | 1028 | 740 |
| **Characteristics of Takotsubo syndrome–related videos (mean)** | | | | | | |
| Duration (min) | 47 | 46 | 89 | 57 | 46 | 45 |
| Likes | 10,048 | 3326 | 163 | 11,861 | 4761 | 82 |
| Comments | 1066 | 229 | 10 | 1002 | 213 | 39 |
| Reposts | 1473 | 294 | 8 | 1113 | 243 | 43 |
| **Video production, n (%)** | | | | | | |
| Presence of people | 20 (91) | 7 (39) | 12 (60) | 9 (82) | 0 (0) | 1 (33) |
| Background music | 12 (55) | 15 (83) | 15 (75) | 8 (73) | 4 (67) | 2 (67) |
| On-screen emoji | 6 (27) | 4 (22) | 2 (10) | 0 (0) | 3 (50) | 0 (0) |
| Animation | 1 (5) | 8 (44) | 5 (25) | 0 (0) | 4 (67) | 0 (0) |
| Subtitles | 21 (96) | 18 (100) | 17 (85) | 9 (82) | 5 (83) | 2 (67) |

## Video Content

The average total score of Goobie et al's [34] 6-question survey by both raters was 5.40 out of 12, indicating that the overall content quality of these videos was average (in our study, the average total score of Goobie et al's [34] 6-question survey ≥9.0 is considered excellent, ≥6.0-<9.0 is considered good, and <6.0 is considered average). The average scores for the videos given by the 2 raters for each of the 6 criteria described by Goobie et al [34] are shown in Multimedia Appendix 2. Of the 6 items, "symptoms" was the most common, while "diagnosis" was the least common. The average scores (total score for each item is 2), graded for each item and ranked from highest to lowest were as follows: symptoms (1.42), risk factors (1.09), definition (1.06), outcome (0.83), management (0.66), and diagnosis (0.34). We further analyzed video authors by categorizing them into the professional and nonprofessional group. Health professionals comprised the professional group, while the nonprofessional group included authors other than health professionals, such as news agencies and for-profit organizations. Health professionals scored higher in terms of video content compared to the other 5 author categories (Table 3). Videos posted by professionals (6.20) scored higher on average than those posted by nonprofessionals (5.09).

**Table 3.** Scoring on various instruments for videos related to Takotsubo syndrome categorized by sources.

| | Health professionals | Individual science communicators | General users | News agencies | For-profit organizations | Nonprofit organizations | *P* value[a] |
|---|---|---|---|---|---|---|---|
| Six questions | 6.20 | 4.81 | 4.65 | 5.77 | 6.08 | 5.33 | <.001 |
| DISCERN score | 40.59 | 36.78 | 30.80 | 38.36 | 40.33 | 39.67 | <.001 |
| JAMA[b] criteria | 1.91 | 1.08 | 1.08 | 1.77 | 1.83 | 2.00 | <.001 |

[a] *P* value obtained by performing the Kruskal-Wallis test; null hypothesis: no difference among the average scores for the 6 groups. It assumes that the 2 coders performed a homogeneous assessment

[b] JAMA: Journal of the American Medical Association.

## Video Quality

The average score given for the videos on the DISCERN instrument by both raters was 36.93 out of 80, indicating that the overall quality of these videos was poor (average total score of 16-26 is very poor, a score of 27-38 is poor, a score of 39-50 is fair, a score of 51-62 is good, and a score >63 is excellent [36]). Among videos created by different author types, the average score given to videos made by health professionals was the highest (40.59), followed by for-profit organizations (40.33) and non-profit organizations (39.67). Although the video scores of the above 3 resources are relatively higher, the quality rating level is only fair. The lowest average score was for those created by general users (30.80), indicating that videos posted by general users were the poorest in terms of quality. The comparison of scores between the professional and nonprofessional group is shown in Multimedia Appendix 3. Questions 3, 6, and 8 were scored above 3 points, and these questions addressed relevance, bias, and areas of uncertainty with reference to the selected videos, respectively. Questions 5 and 9-14 were found to have scores of less than 2 points. These questions mainly addressed whether the date of the information used or reported in the publication was clear; whether the mechanism, benefits, and risks of each treatment, consequences of not treating, or impact of different treatment options on the overall quality of life were described; and whether the possibility of more than one treatment choice being available was clear (Multimedia Appendix 3).

Both raters agreed that 98% (78/80) of the videos provided the latest information, 43% (34/80) of the videos provided information regarding authorship, but none mentioned the disclosure statement. None of the videos met all the JAMA criteria. The average JAMA score was 1.51. By video author categories, an average score of 1.91 was obtained by health professionals, 1.08 by individual science communicators, 1.08 by general users, 1.77 by news agencies, 1.83 by for-profit organizations, and 2.00 by nonprofit organizations. Based on the Kruskal-Wallis test, there were statistically significant differences in content, DISCERN instrument scores, and the JAMA standard video assessment scores among the 6 author groups ($P$<.05).

## Analysis of Information Quality and Audience Engagement

The results of our multiple linear regression analysis showed that author types that identified as health professionals (exponentiated regression coefficient 17.48, 95% CI 2.29-133.52; $P$=.006) and individual science communicators (exponentiated regression coefficient 13.38, 95% CI 1.83-97.88; $P$=.01) were significant and independent determinants of audience engagement (in terms of the number of likes). Other author types of videos, video content, and DISCERN document scores were not associated with higher likes (Table 4).

**Table 4.** Multiple linear regression analysis on audience engagement[a].

| Variable | Log-transformed data on audience engagement | | | | Original audience engagement | | *P* value |
|---|---|---|---|---|---|---|---|
| | Estimated intercept | SE | *t (df)* | 95% CI | Estimated intercept | 95% CI | |
| **Author types** | | | | | | | |
| Intercept | 4.1848 | 2.5466 | 1.643 (70) | –0.89 to 9.26 | 65.68 | 0.41 to 10550.22 | .11 |
| Health professionals | 2.8609 | 1.0195 | 2.806 (70) | 0.83 to 4.89 | 17.48 | 2.29 to 133.52 | .006 |
| Individual science communicators | 2.5937 | 0.9978 | 2.599 (70) | 0.60 to 4.58 | 13.38 | 1.83 to 97.88 | .01 |
| News agencies | 0.638 | 1.0957 | 0.582 (70) | –1.55 to 2.82 | 1.89 | 0.21 to 16.83 | .56 |
| For-profit organizations | 1.8838 | 1.3658 | 1.379 (70) | –0.84 to 4.61 | 6.58 | 0.43 to 100.26 | .17 |
| Nonprofit organizations | 0.8744 | 1.8097 | 0.483 (70) | –2.73 to 4.48 | 2.40 | 0.06 to 88.56 | .63 |
| Video content[b] | 0.4602 | 0.3158 | 1.457 (70) | –0.17 to 1.09 | 1.58 | 0.84 to 2.97 | .15 |
| **DISCERN instrument** | | | | | | | |
| Reliability of the videos (items 1-8) | –0.6326 | 1.3684 | –0.462 (70) | –3.36 to 2.10 | 0.53 | 0.03 to 8.14 | .65 |
| Quality of treatment choices (items 9-15) | –1.1136 | 1.3283 | –0.838 (70) | –3.76 to 1.54 | 0.33 | 0.02 to 4.64 | .41 |
| Overall information quality (item 16) | 0.3339 | 0.8090 | 0.413 (70) | –1.28 to 1.95 | 1.40 | 0.28 to 7.01 | .68 |

[a] Residual standard error: 2.632 on 70 degrees of freedom; multiple $R^2$=0.2048; adjusted $R^2$=0.1025; $F_{9,70}$=2.003; $P$=.05.

[b] Six questions obtained from reports of Goobie et al [34] were used to measure the video content.

## Discussion

### Principal Findings

Growing evidence suggests that TTS is a severe cardiac disorder with a substantial mortality risk. A meta-analysis showed no difference in the in-hospital and long-term mortality between TTS and ACS [37]. Long-term recurrence rates of TTS have been reported to be as high as 11.4% over 4 years [38]. Although regional wall dysfunction is reversible in TTS, patients may continue to experience chest pain, fatigue, and dyspnea even after recovery of wall function [39]. Notably, traditional cardiovascular risk factors are less commonly seen in TTS compared with those seen in ACS [37]. TTS presentations are often instigated by stress-related emotional and physical factors preceding symptoms [1]. Although such factors may not always be preventable, awareness of TTS as an entity by the general population may improve the measures adopted by individuals in its recognition and seeking prompt medical attention. Additionally, it may improve the collateral history provided by affected individuals, which may aid health care providers in its diagnosis. This is especially important as patients who develop TTS are known to be more apathetic with regard to their mortality and their acute presentation, as evidenced by a study [40] evaluating the psychology of patients with TTS and ACS. That study [40] has also shown that patients with TTS were less likely to be concerned about contracting diseases ($P<.05$) and had fewer thoughts related to the acute cardiac episode that interfered with their life ($P<.001$). It is still unclear whether awareness of TTS will change these perceptions, but an increase in awareness may certainly encourage individuals to adopt measures to mitigate individual stressors, which, in turn, may reduce the incidence of TTS. Some patients with TTS continue to experience symptoms despite resolution of the acute phase. Cardiac rehabilitation may be beneficial for these patients to improve the quality of life and reduce episodes of ongoing chest pain, but only few patients have been reported to receive cardiac rehabilitation owing to the lack of awareness among patients and even doctors [41]. Improving the awareness of TTS may aid the rate of those seeking cardiac rehabilitation for persistent symptoms as well. There is an increasing trend toward using social media for patient education. Despite there being a lot of relevant information on social media platforms from reliable sources, there also exists a large amount of inaccurate information. This poses a significant challenge to users seeking easily accessible and comprehensible information regarding their health. Moreover, low-quality information weakens the ability of people to make informed decisions and can even lead to harmful consequences.

### Video Quality Analysis

To the best of our knowledge, this is the first study to analyze the quality of TTS-related videos on TikTok. As a popular social media platform, TikTok has the potential to become a valuable and influential platform to disseminate health information, especially in the context of the current COVID-19 pandemic [42,43]. In our study, 80 TTS-related videos received 443,469 likes and were commented on and shared thousands of times, which also affirms TikTok's powerful communication capabilities. However, there is significant concern regarding the quality of these TTS videos on TikTok. Their scores on the DISCERN instrument (36.93) and the JAMA criteria (1.51) were generally low. These findings are similar to the results of Śledzińska et al [44] in their study on YouTube videos (n=61) on meningioma treatment. In their study, the mean total DISCERN score was poor as well at 36.4. Part of the reason for the poor quality may be that short-video platforms lack scientific insight for health information dissemination. Furthermore, there is no restriction on the type of content that video authors may publish nor is there any restriction based on author type to ensure content quality. Addressing these concerns in the context of health information dissemination will be invaluable [43]. Considering this, we further studied the quality of videos based on author type. We found that health professionals are the leading creators of popular science videos (Multimedia Appendix 1). Videos created by health professionals had higher average DISCERN scores than those created by non–health professionals. This is consistent with current literature reports that videos created by professionals are likely to have higher quality [29,45,46]. Given the inconsistency in the quality of videos based on the source, we recommend that patients be cautious when obtaining health-related information through platforms such as TikTok.

Notably, videos created by health professionals also received the highest number of likes and reposts in our study, which is in contrast to the results of other prior studies. In previous studies evaluating videos on psoriasis and nephrolithiasis on YouTube, poor-quality videos received greater attention [19,47]. Our study shows that higher health professional participation in health-related short video productions results in higher popularization ranks, having gained wider user attention. Although the quality of the videos produced by health professionals needs to be improved, it is undeniable that the participation of medical professionals in the creation of high-quality videos plays an essential role in promoting health education [29].

### Content and Optimization Analysis

In terms of video content, we found that very few videos introduced the concept of diagnosis of TTS, and the display of reference information sources was not common, which remains an area to be addressed. The comprehensiveness and accuracy of the video content is a necessary prerequisite to ensure reliable transmission of information, especially when most TikTok users do not have the ability to differentiate health information for reliability. Therefore, we recommend that video producers provide sources of reference information. In addition, in the process of screening videos, we noticed that many videos showed the causes of TTS to be mainly emotional factors, and only a few videos emphasized physical factors, which are important and cannot be ignored. A retrospective study on patients with TTS showed that physical factors are considerable risk factors for in-patient mortality [48]. It is appropriate for video producers to ask experts for review before content creation or check authoritative source materials to ensure that the information published is comprehensive and accurate. As put forward by Oh and You [49], it may be essential to form an expert evaluation team to authenticate the reliability of health-related videos prior to dissemination and provide corresponding identification certificates. In addition, organizational formats of health information can affect the health intervention. Therefore, video authors should fully understand individuals' needs regarding health information and organize information effectively to provide targeted health information [50]. The video monitoring platform should also design a recommendation algorithm that filters higher-quality videos for users as much as possible, especially with reference to medical- and health-related information.

### Audience Engagement Analysis

Through multiple linear regression analysis, we found that, compared with videos made by general authors, those made by health professionals and individual science communicators were more likely to obtain likes. Interestingly, no correlation was found between audience engagement and a video's DISCERN scores. The number of likes is commonly viewed as a collective filter and as an indicator of popularity, which may reflect video quality [19,51,52]. However, our findings suggested that, similar to that in other platforms, the audience on the TikTok platform pay more attention to the identity of the author rather than the content and quality of the video. This finding echoes with those of previous studies [16,19], which show that consumers should remain cautious of using such indictors to judge a video's credibility and that health professionals and individual science communicators have a significant role in video production in TikTok. By providing high-quality videos through these professional authors, TikTok can allow for accelerated health care information dissemination and even potentially improve outcomes in certain diseases.

### Limitations and Future Directions

Our study has some limitations. First, this was a cross-sectional study with a small sample size, despite attempts to include all relevant videos. Second, we only evaluated Chinese videos, which may not be representative in a global context. Future research studies can target characteristics of videos in various languages and regions. In addition, we utilized common standards that are currently used for health information video evaluation, namely, the DISCERN instrument and the JAMA criteria [30,34]. However, some scholars have suggested that these 2 standards were developed relatively early and were initially used to evaluate website information and may have limitations when used to evaluate video information [53]. It is vital that a new video information evaluation tool is developed to adapt in this era of rapidly proliferating video content. Finally, we did not analyze the user comments addressed in each video, and we were unable to track the behavioral and psychological changes of the end user or the information recipient. How these videos affect patient behavior in reality is an area that needs to be studied in the future.

### Conclusion

By analyzing the quality of TTS-related videos on TikTok, we found that videos produced by health professionals were found to have the highest DISCERN scores. However, the overall quality of the videos related to TTS on TikTok is poor. The multiple linear regression analysis showed that author type categories of health professionals (*P*=.006) and individual science communicators (*P*=.01) were significant and independent determinants of consumer engagement (in terms of the number of likes). Our study indicates that the formulation of a measure to review video quality and reliability, especially with respect to health-related information dissemination on TikTok platform, is imperative and patients should be cautious when obtaining health-related information through TikTok. Medical professionals and individual science communicators should be encouraged to create high-quality health-related videos, which may potentially have higher audience engagement and promote health education.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Video authorship of Takotsubo syndrome–related videos.
[PNG File , 102 KB - jmir_v24i9e39360_app1.png ]

Multimedia Appendix 2
Takotsubo syndrome–related videos on TikTok scored per criteria described by Goobie et al [34].
[PNG File , 12 KB - jmir_v24i9e39360_app2.png ]

Multimedia Appendix 3
The average DISCERN score per question categorized by groups.
[PNG File , 31 KB - jmir_v24i9e39360_app3.png ]

## References

1. Ghadri JR, Cammann VL, Jurisic S, Seifert B, Napp LC, Diekmann J, InterTAK co-investigators. A novel clinical score (InterTAK Diagnostic Score) to differentiate takotsubo syndrome from acute coronary syndrome: results from the International Takotsubo Registry. Eur J Heart Fail 2017 Aug;19(8):1036-1042 [FREE Full text] [doi: 10.1002/ejhf.683] [Medline: 27928880]

2. Ghadri J, Wittstein IS, Prasad A, Sharkey S, Dote K, Akashi YJ, et al. International Expert Consensus Document on Takotsubo Syndrome (Part I): Clinical Characteristics, Diagnostic Criteria, and Pathophysiology. Eur Heart J 2018 Jun 07;39(22):2032-2046 [FREE Full text] [doi: 10.1093/eurheartj/ehy076] [Medline: 29850871]
3. Shao Y, Redfors B, Lyon AR, Rosengren A, Swedberg K, Omerovic E. Trends in publications on stress-induced cardiomyopathy. Int J Cardiol 2012 Jun 14;157(3):435-436. [doi: 10.1016/j.ijcard.2012.03.166] [Medline: 22521379]
4. Murugiah K, Wang Y, Desai NR, Spatz ES, Nuti SV, Dreyer RP, et al. Trends in Short- and Long-Term Outcomes for Takotsubo Cardiomyopathy Among Medicare Fee-for-Service Beneficiaries, 2007 to 2012. JACC Heart Fail 2016 Mar;4(3):197-205 [FREE Full text] [doi: 10.1016/j.jchf.2015.09.013] [Medline: 26746377]
5. Prasad A, Lerman A, Rihal CS. Apical ballooning syndrome (Tako-Tsubo or stress cardiomyopathy): a mimic of acute myocardial infarction. Am Heart J 2008 Mar;155(3):408-417. [doi: 10.1016/j.ahj.2007.11.008] [Medline: 18294473]
6. Templin C, Ghadri JR, Diekmann J, Napp LC, Bataiosu DR, Jaguszewski M, et al. Clinical Features and Outcomes of Takotsubo (Stress) Cardiomyopathy. N Engl J Med 2015 Sep 03;373(10):929-938. [doi: 10.1056/NEJMoa1406761] [Medline: 26332547]
7. Pillière R, Mansencal N, Digne F, Lacombe P, Joseph T, Dubourg O. Prevalence of tako-tsubo syndrome in a large urban agglomeration. Am J Cardiol 2006 Sep 01;98(5):662-665. [doi: 10.1016/j.amjcard.2006.03.048] [Medline: 16923457]
8. Azzarelli S, Galassi AR, Amico F, Giacoppo M, Argentino V, Tomasello SD, et al. Clinical features of transient left ventricular apical ballooning. Am J Cardiol 2006 Nov 01;98(9):1273-1276. [doi: 10.1016/j.amjcard.2006.05.065] [Medline: 17056345]
9. Stiermaier T, Moeller C, Oehler K, Desch S, Graf T, Eitel C, et al. Long-term excess mortality in takotsubo cardiomyopathy: predictors, causes and clinical consequences. Eur J Heart Fail 2016 Jun;18(6):650-656 [FREE Full text] [doi: 10.1002/ejhf.494] [Medline: 26990821]
10. Jiang S, Liu J. Examining the relationship between Internet health information seeking and patient-centered communication in China: taking into account self-efficacy in medical decision-making. Chinese Journal of Communication 2020 May 25;13(4):407-424. [doi: 10.1080/17544750.2020.1769700]
11. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
12. Kong W, Song S, Zhao YC, Zhu Q, Sha L. TikTok as a Health Information Source: Assessment of the Quality of Information in Diabetes-Related Videos. J Med Internet Res 2021 Sep 01;23(9):e30409 [FREE Full text] [doi: 10.2196/30409] [Medline: 34468327]
13. Ouyang Y. Student's death highlights gaps in China's health regulations. The Lancet Oncology 2016 Jun;17(6):709. [doi: 10.1016/s1470-2045(16)30164-4]
14. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
15. Szmuda T, Syed MT, Singh A, Ali S, Özdemir C, Słoniewski P. YouTube as a source of patient information for Coronavirus Disease (COVID-19): A content-quality and audience engagement analysis. Rev Med Virol 2020 Sep;30(5):e2132 [FREE Full text] [doi: 10.1002/rmv.2132] [Medline: 32537771]
16. Szmuda T, Alkhater A, Albrahim M, Alquraya E, Ali S, Dunquwah RA, et al. YouTube as a source of patient information for stroke: A content-quality and an audience engagement analysis. J Stroke Cerebrovasc Dis 2020 Sep;29(9):105065 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2020.105065] [Medline: 32807469]
17. Sedgewick JA, Arnold EP, Stamm MA, Mulcahey MK. Orthopaedic Sports Medicine Podcasts Should Tailor Characteristics Such as Episode Length and Social Media Utilization for Best Audience Engagement. Arthrosc Sports Med Rehabil 2022 Jun;4(3):e957-e964 [FREE Full text] [doi: 10.1016/j.asmr.2022.01.008] [Medline: 35747629]
18. Szmuda T, Özdemir C, Fedorow K, Ali S, Słoniewski P. YouTube as a source of information for narcolepsy: A content-quality and optimization analysis. J Sleep Res 2021 Apr;30(2):e13053 [FREE Full text] [doi: 10.1111/jsr.13053] [Medline: 32315117]
19. Huang MM, Winoker JS, Allaf ME, Matlaga BR, Koo K. Evidence-based quality and accuracy of YouTube videos about nephrolithiasis. BJU Int 2021 Feb;127(2):247-253. [doi: 10.1111/bju.15213] [Medline: 32805761]
20. Chen Q, Min C, Zhang W, Ma X, Evans R. Factors Driving Citizen Engagement With Government TikTok Accounts During the COVID-19 Pandemic: Model Development and Analysis. J Med Internet Res 2021 Feb 04;23(2):e21463 [FREE Full text] [doi: 10.2196/21463] [Medline: 33481756]
21. Song S, Zhao YC, Yao X, Ba Z, Zhu Q. Short video apps as a health information source: an investigation of affordances, user experience and users’ intention to continue the use of TikTok. INTR 2021 Jun 24;31(6):2120-2142. [doi: 10.1108/intr-10-2020-0593]
22. Song S, Zhao YC, Yao X, Ba Z, Zhu Q. Serious information in hedonic social applications: affordances, self-determination and health information adoption in TikTok. JD 2021 Nov 19;78(4):890-911. [doi: 10.1108/jd-08-2021-0158]
23. Stokel-Walker C. TikTok wants longer videos-whether you like it or not. Wired.com. URL: https://www.wired.com/story/tiktok-wants-longer-videos-like-not/ [accessed 2022-07-10]

24. Lovett JT, Munawar K, Mohammed S, Prabhu V. Radiology Content on TikTok: Current Use of a Novel Video-Based Social Media Platform and Opportunities for Radiology. Curr Probl Diagn Radiol 2021;50(2):126-131. [doi: 10.1067/j.cpradiol.2020.10.004] [Medline: 33250298]
25. Zheng DX, Mulligan KM, Scott JF. TikTok and dermatology: An opportunity for public health engagement. J Am Acad Dermatol 2021 Jul;85(1):e25-e26. [doi: 10.1016/j.jaad.2021.02.050] [Medline: 33639245]
26. Basch CH, Mohlman J, Fera J, Tang H, Pellicane A, Basch CE. Community Mitigation of COVID-19 and Portrayal of Testing on TikTok: Descriptive Study. JMIR Public Health Surveill 2021 Jun 10;7(6):e29528. [doi: 10.2196/29528]
27. Chen MKY, Garden F, Sebaratnam DF. Isotretinoin on TikTok™: misinformation getting under our skin. Clin Exp Dermatol 2021 Dec;46(8):1606-1607. [doi: 10.1111/ced.14810] [Medline: 34155690]
28. Song S, Xue X, Zhao YC, Li J, Zhu Q, Zhao M. Short-Video Apps as a Health Information Source for Chronic Obstructive Pulmonary Disease: Information Quality Assessment of TikTok Videos. J Med Internet Res 2021 Dec 20;23(12):e28318 [FREE Full text] [doi: 10.2196/28318] [Medline: 34931996]
29. Om A, Ijeoma B, Kebede S, Losken A. Analyzing the Quality of Aesthetic Surgery Procedure Videos on TikTok. Aesthet Surg J 2021 Nov 12;41(12):2078-2083. [doi: 10.1093/asj/sjab291] [Medline: 34297053]
30. Eksi Ozsoy H. Evaluation of YouTube videos about smile design using the DISCERN tool and Journal of the American Medical Association benchmarks. J Prosthet Dent 2021 Jan;125(1):151-154. [doi: 10.1016/j.prosdent.2019.12.016] [Medline: 32085870]
31. Stern J, Georgsson S, Carlsson T. Quality of web-based information at the beginning of a global pandemic: a cross-sectional infodemiology study investigating preventive measures and self care methods of the coronavirus disease 2019. BMC Public Health 2021 Jun 14;21(1):1141 [FREE Full text] [doi: 10.1186/s12889-021-11141-9] [Medline: 34126962]
32. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
33. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
34. Goobie GC, Guler SA, Johannson KA, Fisher JH, Ryerson CJ. YouTube Videos as a Source of Misinformation on Idiopathic Pulmonary Fibrosis. Ann Am Thorac Soc 2019 May;16(5):572-579. [doi: 10.1513/AnnalsATS.201809-644OC] [Medline: 30608877]
35. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
36. Olkun HK, Ari Demirkaya A. Evaluation of Internet Information about Lingual Orthodontics Using DISCERN and JAMA Tools. Turk J Orthod 2018 Jun 01;31(2):50-54. [doi: 10.5152/turkjorthod.2018.17042]
37. Han P, Yang Z, Diao K, Huang S, Shen M, Zhang Y, et al. Comparison of clinical profiles between takotsubo syndrome and acute coronary syndrome: a systematic review and meta-analysis. Heart Fail Rev 2020 Sep;25(5):847-860. [doi: 10.1007/s10741-019-09846-6] [Medline: 31446536]
38. Elesber AA, Prasad A, Lennon RJ, Wright RS, Lerman A, Rihal CS. Four-year recurrence rate and prognosis of the apical ballooning syndrome. J Am Coll Cardiol 2007 Jul 31;50(5):448-452 [FREE Full text] [doi: 10.1016/j.jacc.2007.03.050] [Medline: 17662398]
39. Singh K. Tako-Tsubo syndrome: issue of incomplete recovery and recurrence. Eur J Heart Fail 2016 Dec;18(12):1408-1410 [FREE Full text] [doi: 10.1002/ejhf.574] [Medline: 27221009]
40. Gorini A, Galli F, Giuliani M, Pierobon A, Werba JP, Adriano EP, et al. Psychological Characteristics of Patients with Takotsubo Syndrome and Patients with Acute Coronary Syndrome: An Explorative Study toward a Better Personalized Care. J Pers Med 2022 Jan 04;12(1):38 [FREE Full text] [doi: 10.3390/jpm12010038] [Medline: 35055353]
41. Wu CM, McKeon J, Abbott JD, Jiang L, Wu W. Referral to Cardiac Rehabilitation and Outcomes for Patients With Takotsubo Cardiomyopathy. J Cardiopulm Rehabil Prev 2019 May;39(3):E8-E11. [doi: 10.1097/HCR.0000000000000433] [Medline: 31022006]
42. Basch CH, Hillyer GC, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 2020 Aug 10:2020. [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
43. Eghtesadi M, Florea A. Facebook, Instagram, Reddit and TikTok: a proposal for health authorities to integrate popular social media platforms in contingency planning amid a global pandemic outbreak. Can J Public Health 2020 Jun;111(3):389-391 [FREE Full text] [doi: 10.17269/s41997-020-00343-0] [Medline: 32519085]
44. Śledzińska P, Bebyn MG, Furtak J. Quality of YouTube Videos on Meningioma Treatment Using the DISCERN Instrument. World Neurosurg 2021 Sep;153:e179-e186. [doi: 10.1016/j.wneu.2021.06.072] [Medline: 34166826]
45. Villa-Ruiz C, Kassamali B, Mazori DR, Min M, Cobos G, LaChance A. Overview of TikTok's most viewed dermatologic content and assessment of its reliability. J Am Acad Dermatol 2021 Jul;85(1):273-274. [doi: 10.1016/j.jaad.2020.12.028] [Medline: 33359080]

46. Stellefson M, Chaney B, Ochipa K, Chaney D, Haider Z, Hanik B, et al. YouTube as a source of chronic obstructive pulmonary disease patient education: a social media content analysis. Chron Respir Dis 2014 May;11(2):61-71 [FREE Full text] [doi: 10.1177/1479972314525058] [Medline: 24659212]
47. Mueller SM, Jungo P, Cajacob L, Schwegler S, Itin P, Brandt O. The Absence of Evidence is Evidence of Non-Sense: Cross-Sectional Study on the Quality of Psoriasis-Related Videos on YouTube and Their Reception by Health Seekers. J Med Internet Res 2019 Jan 16;21(1):e11935 [FREE Full text] [doi: 10.2196/11935] [Medline: 30664460]
48. Sobue Y, Watanabe E, Ichikawa T, Koshikawa M, Yamamoto M, Harada M, et al. Physically triggered Takotsubo cardiomyopathy has a higher in-hospital mortality rate. Int J Cardiol 2017 May 15;235:87-93. [doi: 10.1016/j.ijcard.2017.02.090] [Medline: 28262347]
49. Oh J, You SY. Febrile seizure: What information can caregivers access through YouTube? Seizure 2021 Oct;91:91-96. [doi: 10.1016/j.seizure.2021.05.020] [Medline: 34130196]
50. Xu X, Yang M, Zhao YC, Zhu Q. Effects of message framing and evidence type on health information behavior: the case of promoting HPV vaccination. AJIM 2020 Dec 07;73(1):63-79. [doi: 10.1108/ajim-02-2020-0055]
51. Yang S, Brossard D, Scheufele DA, Xenos MA. The science of YouTube: What factors influence user engagement with online science videos? PLoS One 2022;17(5):e0267697 [FREE Full text] [doi: 10.1371/journal.pone.0267697] [Medline: 35613095]
52. Erdem MN, Karaca S. Evaluating the Accuracy and Quality of the Information in Kyphosis Videos Shared on YouTube. Spine (Phila Pa 1976) 2018 Nov 15;43(22):E1334-E1339. [doi: 10.1097/BRS.0000000000002691] [Medline: 29664816]
53. Azer SA. Are DISCERN and JAMA Suitable Instruments for Assessing YouTube Videos on Thyroid Cancer? Methodological Concerns. J Cancer Educ 2020 Dec;35(6):1267-1277. [doi: 10.1007/s13187-020-01763-9] [Medline: 32472374]

## Abbreviations

**ACS:** acute coronary syndrome
**JAMA:** Journal of the American Medical Association
**TTS:** Takotsubo syndrome

*Edited by G Eysenbach; submitted 09.05.22; peer-reviewed by X Xu, Q Chen; comments to author 17.06.22; revised version received 15.07.22; accepted 30.08.22; published 26.09.22.*

*Please cite as:*
*Liang J, Wang L, Song S, Dong M, Xu Y, Zuo X, Zhang J, Adrian Sherif A, Ehsan J, Ma J, Li P*
*Quality and Audience Engagement of Takotsubo Syndrome–Related Videos on TikTok: Content Analysis*
*J Med Internet Res 2022;24(9):e39360*
*URL: https://www.jmir.org/2022/9/e39360*
*doi:10.2196/39360*
*PMID:36155486*

Original Paper

# Exploring Nurse and Patient Experiences of Developing Rapport During Oncology Ambulatory Care Videoconferencing Visits: Qualitative Descriptive Study

Paula D Koppel[1*], MS, RNC, GNP-BC, AHN-BC, NBC-HWC; Jennie C De Gagne[1*], DNP, RN, NPD-BC, CNE, PhD; Sharron Docherty[1*], PNP, RN, PhD; Sophia Smith[1,2*], MSW, PhD; Neil S Prose[3,4*], MD; Terri Jabaley[5*], RN, OCN, PhD

[1]School of Nursing, Duke University, Durham, NC, United States
[2]Duke Cancer Institute, Duke Health, Durham, NC, United States
[3]Department of Dermatology, Duke University, Durham, NC, United States
[4]Department of Pediatrics, Duke University, Durham, NC, United States
[5]Phyllis F Cantor Center for Research in Nursing & Patient Care Services, Dana-Farber Cancer Institute, Boston, MA, United States
[*]all authors contributed equally

**Corresponding Author:**
Paula D Koppel, MS, RNC, GNP-BC, AHN-BC, NBC-HWC
School of Nursing
Duke University
DUMC 3322
307 Trent Drive
Durham, NC, 27710
United States
Phone: 1 617 835 7087
Fax: 1 919 681 8899
Email: paula.koppel@duke.edu

## Abstract

**Background:** Although videoconferencing between oncology patients and nurses became routine during the pandemic, little is known about the development of clinician-patient rapport in this care environment. Evidence that virtual visits may challenge nurses' ability to form connections with patients, demonstrate empathy, and provide support suggests that videoconferencing may not ensure optimal care for persons with cancer. Establishing rapport during videoconferencing visits (VCVs) is important in oncology nursing, as rapport enables the nurse to provide emotional support and assistance to patients as they navigate their cancer journey.

**Objective:** This study investigated the nature of nurse-patient rapport in ambulatory cancer care videoconferencing telehealth visits. Objectives included exploring (1) how patients with cancer and nurses describe experiences of and strategies for cultivating rapport and (2) similarities and differences between rapport in videoconferencing and in-person visits (IPVs).

**Methods:** In this qualitative descriptive study, interviews were conducted from October 2021 to March 2022 with 22 participants, including patients with cancer (n=10, 45%) and oncology nurses (n=12, 55%), about their experiences of rapport building during VCVs. All interviews were analyzed using conventional content analysis. Data from nurses and patients were analyzed separately using identical procedures, with a comparative analysis of patient and nurse results performed in the final analysis.

**Results:** Most patients in the study had experienced 3-5 video visits within the past 12 months (n=7, 70%). Half of the nurse participants (n=6, 50%) reported having participated in over 100 VCVs, and all had experiences with videoconferencing (ranging from 3 to 960 visits) over the past 12 months. In total, 3 themes and 6 categories were derived from the patient data, and 4 themes and 13 categories were derived from the nurse data. Comparisons of themes derived from participant interviews identified similarities in how nurses and patients described experiences of rapport during VCVs. Three themes fit the collective data: (1) person-centered and relationship-based care is valued and foundational to nurse-patient rapport in oncology ambulatory care regardless of how care is delivered, (2) adapting a bedside manner to facilitate rapport during VCVs is feasible, and (3) nurses and patients can work together to create person-centered options across the care trajectory to ensure quality care outcomes. Barriers to relationship building in VCVs included unexpected interruptions from others, breaks in the internet connection, concerns about privacy, and limitations associated with not being physically present.

**Conclusions:** Person-centered and relationship-based approaches can be adapted to support nurse-patient rapport in VCVs, including forming a personal connection with the patient and using active listening techniques. Balancing the challenges and limitations with the benefits of videoconferencing is an essential competency requiring additional research and guidelines.

**International Registered Report Identifier (IRRID):** RR2-10.2196/27940

***(J Med Internet Res 2022;24(9):e39920)*** doi:10.2196/39920

**KEYWORDS**

clinician-patient relationship; nursing; oncology ambulatory care; rapport; telehealth; videoconferencing visits; videoconferencing; telemedicine; ambulatory care; cancer care; oncology nurse

# *Introduction*

## Background

During the COVID-19 global pandemic, the use of telehealth to provide oncology care as safely as possible increased dramatically. Some National Cancer Institute (NCI)–designated cancer centers reported that 33%-50% of all cancer patient encounters were telehealth visits, with utilization increases as high as 4693% compared to prepandemic rates [1]. Although telehealth includes a variety of technologies that support long-distance clinical health care, the use of videoconferencing visits (VCVs) as a modality to care for patients during the pandemic rapidly expanded in oncology during the pandemic [2]. Despite initial logistical challenges, many patients and oncology health care clinicians have decided that videoconferencing provides an innovative and effective way to receive and deliver care, yet little is known about how this modality of care affects patient-clinician relationships.

Review studies of videoconferencing in telehealth indicate that this computer-mediated modality has utility and comparable outcomes to in-person visits (IPVs) for patients with various chronic conditions [3-6]. Systematic reviews in oncology suggest that VCVs are feasible and can be effective for certain types of cancer care [7,8]. Prepandemic studies focused on videoconferencing for palliative care consultation and provision of support between patients, family caregivers, and community-based care clinicians [9-11] indicated that VCVs are often preferable for palliative care consultations [11,12], hospice family meetings [13,14], and support groups [15]. Recent studies suggest that patients are similarly satisfied with cancer care received via videoconferencing [16-18] and that oncology clinicians are increasingly receptive to videoconferencing and acknowledge its numerous benefits to patients [2].

## Knowledge Gaps

Although videoconferencing is feasible and can be effective for certain types of care, nurses, doctors, and mental health clinicians across health disciplines have expressed concern that the two-dimensional nature of videoconferencing interactions, including the loss of physical proximity, presence, and touch, depersonalizes and inhibits care and the clinician’s ability to understand the patient fully [19-22]. In palliative care studies, clinicians have indicated a reluctance to initiate emotional topics during VCVs as they cannot be physically present to provide necessary support [23,24] or ensure that patients have adequate privacy [14]. Suggestions for adapting in-person rapport techniques, such as small talk, eye contact, and body language, for VCVs have been discussed in the popular press and clinical commentaries [25-28]; however, little research has evaluated these modalities [29-32] or more advanced relational skills for providing presence [33-35], conveying caring [36] or empathy [37], and delivering person-centered care [38].

## Importance of Nurse-Patient Rapport

Rapport can be defined as a connection established with another person based on respect, acceptance, empathy, and a mutual commitment to engagement [39,40]. Interpersonal interventions that cultivate rapport between patients and clinicians can potentially improve patient health outcomes and satisfaction [41]. For persons with cancer, feeling personally known and connected with nurses and health care clinicians on a level beyond their disease process reduces suffering and improves satisfaction, health outcomes, and quality of life [42-45]. Research suggests that rapport facilitates a trusting and therapeutic relationship [46,47], enabling the clinician to become a source of emotional support for patients with cancer [39,48-50]. The literature supports that nurse-patient rapport facilitates the trusting relationship necessary to ensure holistic assessment of needs, personalization of care, and adaptive work.

## Critical Need for Research

COVID-19 necessitated a rapid adoption of VCVs into the standard care of patients with cancer, leaving little time to create thoughtful guidelines based on quality improvement or research focused on the impact of VCVs on patient-clinician relationships or care outcomes. Clinicians are seeking guidance regarding ways to transfer interpersonal skills to computer-mediated forms of care [51]. Studies comparing IPVs and VCVs indicate that (1) clinicians use less empathetic, supportive, and facilitating statements [37]; (2) exchange of information is reduced; and (3) patients present fewer problems [52] in VCVs; however, evidence regarding the effect of these findings on rapport and clinical care outcomes is lacking.

## Research Aim and Questions

The purpose of this qualitative descriptive study was to explore the experiences of nurses and patients participating in oncology telehealth VCVs, specifically in relation to the cultivation of rapport. This study explored rapport-building strategies as well as similarities and differences between rapport in VCVs compared to IPVs, with the aim of providing a foundation for relationship-building guidelines for telehealth videoconferencing.

## Methods

### Study Design

This qualitative descriptive design included interviews with patients with cancer and oncology nurses from October 2021 to March 2022. Semistructured interviews were conducted through the secure videoconferencing platform Zoom (Zoom Video Communications Inc). Participants were asked to share their general thoughts and feelings about VCVs; follow-up questions probed more deeply into their experiences of rapport during these visits. This descriptive qualitative approach, as described by Sandelowski [53,54], focused on experiences as directly reported by participants, while limiting deeper interpretation. This study included a sufficient number of nurse and patient participants to describe comprehensively the experiences of nurse-patient rapport in the new context of videoconferencing in ambulatory oncology [53,54], thus providing rich interpersonal experiences of VCVs. A brief description of the study components follows; details are available in the published protocol [55].

### Setting, Recruitment, and Eligibility Criteria

The study was conducted at an NCI-designated comprehensive cancer center in a northeastern metropolitan area of the United States. The scheduling and management of VCVs was centralized and standardized on the oncology center's approved secure Zoom communication platform. As few VCVs occurred in the oncology center prior to COVID-19, in the winter of 2020, nurses were provided with tip sheets on the logistics of virtual visits and a recorded training session that included strategies for completing a physical examination virtually as well as suggestions for enhancing the "webside" manner (the manner in which a clinician interacts with a patient in VCVs) [26].

Purposive sampling was used to recruit participants who had recently used videoconferencing for their ambulatory care visits. Both patients with cancer and oncology nurses were included to enhance the understanding of how rapport building occurs within a nurse-patient dyad. Recruitment for the study began in September 2021. Nurses and patients were recruited remotely through a combination of efforts, including announcements at virtual nursing staff meetings and requests that participating nurses recommend colleagues and patients for the study. Efforts to recruit underrepresented participants were undertaken in the institution's community-based satellite clinics. Patient recruitment focused on individuals who had nurse VCV teaching sessions or had seen their advanced practice nurse within the past 12 months. Participants were not compensated for their participation.

Patient inclusion criteria were (1) adult (18 years or older), (2) able to read and converse in English, (3) currently receiving care at the oncology institute, (4) current or former participant in at least 1 VCV with a nurse from the oncology center, and (5) enrolled in the oncology center's patient portal. Patients with cognitive impairment or any condition that prohibited their ability to provide informed consent (eg, Alzheimer disease or related dementias) were excluded. Nurse inclusion criteria were (1) licensed registered or advanced practice nurse employed at the cancer institute for at least 1 year postorientation and (2) current or former participant in VCVs with patients at the oncology center. No restrictions on the length of nursing practice were necessary, as competency in developing rapport with patients is a foundational skill [56].

### Ethical Considerations

Approval was obtained from the Dana-Farber Cancer Institute Institutional Review Board (IRB) (protocol number: 21-318) and the Duke University Health System IRB for Clinical Investigations (protocol identification number: Pro00108787) prior to beginning the study. Written consent was obtained by the first author (PK) from all participants prior to the collection of any data.

### Data Collection Procedures

All interviews were conducted through the institution-approved platform by the first author (PK), a trained nurse scientist with 3 years of experience in qualitative methods. Although interviews were conducted in a videoconference format, only the audio portion was recorded. Audio recordings were transcribed verbatim. All identifiable participant information was removed, and records were assigned participant IDs to ensure confidential analysis of the text-based data.

Semistructured interviews were conducted. Participants were given an opportunity to tell complete stories about their experiences before the interviewer asked probing follow-up questions from a guide developed by the research team. Similar questions were asked of patient and nurse participants. The interview procedures and guide are published elsewhere [55]. By March 2022, the data collected were large enough to capture the rich experiences of nurses and patients but small enough to permit a thorough analysis [57]. The final interviews identified no new themes related to the research questions, indicating data saturation [58].

### Data Management and Analysis

Conventional content analysis was performed to analyze the narrative qualitative data, given how little is known to date and the consequent need for robust descriptive data. Codes were derived directly from the transcribed text data and kept close to participants' descriptions [53,59]. Nurse and patient interviews were analyzed separately using identical procedures. Elo and Kyngas' [56] content analysis management process was used to organize the analysis process into three phases: (1) preparation, (2) organization, and (3) reporting. This analysis process included creating, defining, and recording codes, categories, and themes and matching themes with exemplar quotations in a codebook using the data analysis management tool NVivo 12.0 (QSR International Pty Ltd). Initially, 2 research team members (authors PK and JDG) independently coded the same cases while compiling the codebook. Once agreement was reached on the codebook and coding of 20% of the transcripts, 1 team member (PK) completed coding the remaining transcripts while meeting with the team to discuss new codes and revise the codebook weekly. After coding 4 transcriptions from each group, the team decided to use 1 codebook to analyze nurse and patient data as the interview questions and responses were similar. Most of the codes were defined to capture both nurse and patient responses, but a small number were specific for each group. Although a common codebook was used, nurse and patient data were initially analyzed separately and then compared to explore similarities and differences. This process led to the discovery of a model to explain the overall research findings. Exemplar quotes from participants provide the evidence for our findings. Both individual and comparative analyses are reported to provide a rich and deep understanding of the data [56,59]. Findings are evaluated within the context of related theories, models, and evidence-based research. The study's rigor, also described as trustworthiness in qualitative research [58,60], was enhanced by (1) conducting all analyses as a team with 18 combined years of qualitative research experience; (2) using weekly coding meetings to discuss and define all codes, categories, and themes; (3) collecting and analyzing the data concurrently [60]; (4) using detailed memos to create an audit trail of analytical decisions [58,61]; (5) confirming categories representing expansive and diverse experiences with exemplar quotations from multiple participants [58,62]; and (6) using member checking techniques by asking participants to clarify questions during interviews [58]. The consolidated criteria for reporting qualitative research (COREQ) checklist guided the reporting of results [63].

## Results

### Participant Characteristics

The NCI-designated comprehensive cancer center and satellite locations from where patients and nurses were recruited has an academic affiliation and Magnet Recognition for nursing excellence; per its primary care model, patients usually interact with the same nurse or advanced practice nurse. During the pandemic, the center used VCVs to follow patients during active treatment, provide education, and facilitate emotional support to online groups. The study sample included 22 participants (10 persons with cancer, 45%; 12 oncology nurses, 55%). Interviews with nurses lasted between 18 and 43 minutes and those with patients from 16 to 48 minutes. Persons with cancer were aged 36-67 years (mean 54.3, SD 8.31). The patient sample included more women (n=7, 70%) than men (n=3, 30%), and all participants identified as White/Caucasian. All patient participants had some college education, with half having graduate or professional degrees (n=5, 50%). Most participants were married (n=7, 70%), employed full-time (n=7, 70%), and had an annual income of US $100,000 or more (n=6, 60%). Patient participants were evenly split between those with a recent diagnosis and treatment of less than 1 year (n=5, 50%) and those having received treatment for over 1 year (n=5, 50%). Most patients had experienced 3-5 VCVs over the past 12 months (n=7, 70%) and at least 1 within the past 3 months (n=7, 70%). All patient participants reported spending at least 1 hour per day on their computers, with the majority spending 4 or more hours per day (n=7, 70%). Half of the participants reported having participated in at least 50 VCVs for work or personal connections within the past year (n=5, 50%). Multimedia Appendix 1 displays detailed characteristics of the patient participants.

Nurse participants (n=12, 55%) were mostly female (n=10, 83%) and identified as White/Caucasian (n=11, 92%). Their average age was 42.25 years (SD 8.62), with most aged either 31-40 years (n=5, 42%) or 41-50 years (n=5, 42%). All had attended at least some college, with the majority having a master's degree (n=8, 67%). Nursing experience within the sample ranged from 4 to 35 years (mean 16.54, SD 8.67). Half of the participants had worked at the cancer center for over 10 years (n=6, 50%). Their experiences with VCVs ranged from less than 15 visits (n=5, 42%) to over 50 visits (n=3, 25%) within the past 3 months. Half of the nurse participants (n=6, 50%) reported having participated in over 100 VCVs within the past 12 months. Multimedia Appendix 2 displays detailed characteristics of the nurse participants.

### Qualitative Analysis of Patient Interviews

The data showed that patients view VCVs with their health care clinicians positively and appreciate having a personal relationship with their nurses, which they find is achievable in videoconferencing. Three key themes were identified during the analysis of patient interviews: (1) building rapport in VCVs and IPVs requires a personal touch, (2) rapport can facilitate trust in VCVs and impact how patients feel about their care, and (3) videoconferencing works well for some visits but is not ideal for others. Multimedia Appendix 3 displays a narrative description of each theme, associated categories, codes, and additional participant quotations.

#### *Theme 1: Building Rapport in Videoconferencing Visits and In-Person Encounters Requires a Personal Touch*

Theme 1 relates to the personal level, focusing on the patient's description of the nurse's (1) understanding of their life and feelings, apart from their diagnosis and treatment and (2) willingness to take time to present information and answer questions.

#### Being Known as a Person, Not a Patient

Some patients expressed that establishing rapport in videoconferencing is harder and that small talk at the beginning of the visit is more important than during IPVs. One participant said:

> *Things that help build the rapport would just be establishing the connection and relationship, like asking those little family questions and taking a moment to get a little bit beyond just the medical piece.* [Patient 9 (P9)]

They appreciated being asked questions about their families, occupation, and things that brought them joy and appreciated the nurses remembering these. One participant said the following:

> *Through conversations, my nurse remembered that my wedding anniversary was in May, my kid's birthday was coming up in July…you build that relationship, and…it makes it that much easier when you go through your treatment.* [P5]

One participant indicated that knowing their nurse on a personal level contributed to their sense that the nurse viewed them as a person, not just a patient. They appreciated the nurse's willingness to share a few things about themselves and glimpses of the nurse's personal environment when videoconferencing was done from the nurse's home.

#### Being Heard and Knowing What Is Important

Participants expressed feeling heard, known, understood, validated, and valued when the nurse used active listening techniques. They had more confidence that the nurse understood what was uniquely important to them. One participant remarked:

> *A big part of rapport is that you want to feel heard and feel responded to in a way that acknowledges that you are heard.* [P6]

#### Taking Time to Provide Information and Answer Questions Thoughtfully

Participants felt the nurse was competent and cared about them as a person when they were given time to ask questions and received comprehensive, personalized answers. One participant noted:

> *I want to…ask as many questions as I need to, [without] a rush on the time…I don't want a pat answer. I want it customized to me or to my situation.* [P6]

### *Theme 2: Rapport Can Facilitate Trust in Videoconferencing and Impact How Patients Feel About Their Care*

This theme (1) focuses on how affect and nonverbal communication impact patients' ability to develop rapport, trust, and confidence with their nurses during VCVs and (2) incorporates the outcomes that participants felt result from rapport, including enhanced comfort and healing.

#### Creating an Open Atmosphere With Positive Affect and Nonverbal Communication

Participants described the importance of the nurse having a friendly demeanor and using eye contact, facial expressions, vocal tones, and other types of body language to convey care, concern, attentiveness, and active listening. Participants felt these behaviors help establish rapport and set the tone in VCVs. One participant stated:

> *Facial cues help establish a good rapport. When you're nodding your head, I realize that you are hearing me, understanding me, and maybe agree or at least sense my viewpoint.* [P6]

Participants felt the nurse's ability to incorporate these behaviors and attributes influences the success of the VCV and encourages open engagement. One participant said:

> *A relationship…has to be two ways. The patient['s]…willing[ness] to engage as well and be open [is] encouraged by the nurse.* [P9]

#### Building Confidence in the Plan of Care

Patients described how rapport and trust affect the sharing of information and build confidence. One patient explained:

> *If I don't have a good relationship with somebody, I don't feel comfortable enough to trust them, to tell them things, be open and honest.* [P9]

Patients described having confidence in their care when they felt their nurse understood what was important to them. One participant stated:

> *If you have a good rapport with them, you know that every decision they're making and everything they're doing is in your best interests.* [P2]

#### Promoting Comfort and Healing

Patients expressed that rapport with their nurses creates a sense of comfort. One participant indicated that rapport contributes to healing:

> *When you feel the [nurse] is looking at you in a compassionate way and caring about you as a human, it contributes to a positive outcome because you're taking in positive energy, and that's how it aids the healing process.* [P6]

One patient was equally enthusiastic about IPVs and VCVs and described in detail a seamless sense of knowing with their nurse:

> *My primary nurse, when she would see me, would have an intuition for how I was doing that particular day, and accommodate quickly, and that's because we'd built a rapport to know that. She could just say one thing to me, and I would know, "Oh, I've got to do this today," because we had established the rapport, and I knew where she was coming from as well, so it allows for ease of the communication and a bidirectional understanding of what the needs are.* [P4]

### *Theme 3: Videoconferencing Works Well for Some Visits but Is Not Ideal for Others*

The third theme focuses on the best circumstances for VCVs and IPVs. All the patient participants (1) expressed that VCVs are beneficial in certain situations, increase convenience, and improve access to clinicians and (2) hoped they would remain available post–COVID-19 for persons with cancer; however, they acknowledged that some oncology care needs to be provided in person.

#### Leveraging the Advantages of Videoconferencing Visits and Technology

Participants indicated that they find videoconferencing convenient and easy and that it saves time, money, and energy. One patient described the ease of reviewing test results with their nurse without disrupting their day:

> *Going over test results with me,…he's showing me the pictures. If you were in person, you'd have to try and peek around the computer…it really is a benefit, and then like 15 minutes are over, and I feel good about myself, and I can go on about my day. I don't need to get in a car and drive somewhere.* [P5]

Participants felt VCVs improve access to their nurses and other health care clinicians.

> *It allows me to go to the site that I want to for the labs and then I can do a video call for the appointment itself. So, it just gives me flexibility [and] more options.* [P10]

One participant indicated their nurse was able to be more attentive during their VCVs:

> *It's more relaxed, and people are under less of a time constraint. At the hospital…people seem more rushed.* [P7]

VCVs reduced the stress associated with travel to appointments, as 1 participant noted:

> *There's so many appointments [for which] you do have to actually be there in person, so to have some things that are taken off that, it really does reduce stress.* [P8]

Patients appreciated the ability to have supportive family members or friends participate in VCVs. Overall, the participants were willing to accept the challenges of a new technology to leverage its advantages.

### Accepting the Challenges of a New Technology

Challenges included learning to use the technology, managing internet connectivity issues, and minimizing distractions. Older participants acknowledged their lack of confidence with videoconferencing technology but seemed empowered by learning to use it. One older participant noted:

> *People my age gets scared by the [idea of] technology…We're getting better…every time it gets easier.* [P7]

Internet connectivity problems were acknowledged as challenges rather than barriers. One participant noted:

> *Losing the connection…That's the way of the world.* [P7]

Participants also mentioned the importance of managing distractions during VCVs. One participant stated:

> *If there's a lot of…distractions on both sides, you're not going to really feel like you're able to talk.* [P3]

All the patients interviewed recognized that some visits are not feasible in videoconferencing. One explained:

> *If you physically have to be there…[for] your cancer treatment, chemo, radiation, anything like that, or…an examination…you can't do some of those things over a video.* [P5]

Others favored an IPV when visualizing physical conditions was important:

> *In person, you can see the struggles physically that the person has. If they're using a wheelchair. If they need assistance to go to the bathroom…but not virtually.* [P1]

Patient participants identified the type of information to be shared as another important determinant of VCV appropriateness. One patient stated:

> *I'm thinking if it was positive news, it's great on Zoom. Negative news might be not so great on Zoom. That might be…a little impersonal.* [P9]

Another participant indicated:

> *I would default to teleconferences over in person if everything else were equal…[an] exception might be if the decision or something you needed to learn was very complex.* [P4]

Another participant commented that they felt less comfortable sharing confidential matters in videoconferencing:

> *If you had something highly personal to discuss, I would feel less comfortable doing it virtually…we don't know how secure it is…[I]f I had problems that I didn't want to say out loud, I'd feel more comfortable saying it with [my] nurse in a little room…somebody else might also be listening.* [P6]

One participant had a difficult time describing why they preferred IPVs:

> *I don't know why [IPVs and VCVs are] different, but I suspect some of it has nothing to do with words or body language or anything [but] just wanting to be in the same room with the person.* [P10]

Of note, this participant had been receiving care over a long period, and videoconferencing meant they lost the support experienced by being in the cancer center and interacting with team members.

Although positive about VCVs, participants recognized the challenges of comprehending complex emotions and situations in their entirety in a virtual environment. They emphasized the importance of combining VCVs and IPVs:

> *You can build a good rapport through Zoom meetings, but it's important that there be face-to-face meetings intermingled. They don't have to be as often, but a combination of the two would be most beneficial.* [P2]

### Finding Minimal Differences in Terms of Communication Effectiveness

Participants largely felt communication with their nurses in VCVs is comparable to that in IPVs. Most of the patients interviewed indicated that they felt comfortable expressing emotions and asking questions in VCVs. One participant expressed:

> *If that rapport is there, if [the nurse and patient are] paying attention to each other, it's just as good as an in-person visit.* [P3]

Another noted the following:

> *I didn't find there was much difference between the two…You could feel the concern. You could feel the genuineness in [the nurse's] voice even though it was on the Zoom, which makes it so much easier to be able to deal with…a fresh diagnosis of cancer.* [P2]

However, some subtle aspects of communication during emotionally charged conversations may be compromised in videoconferencing. As 1 participant explained:

> *Communication [about scan results]…I feel like they're harder to do over video…It's sort of the contextual meaning of the words, if that makes sense…there's going to be a lot of words about the disease, and my question is, "How much time do I have?" That's really at the heart of it, right?... "What does this mean in terms of my quality of life and my quantity of life?" And nobody knows the answer to that, but the doctors and nurses…can give some contextual information. It's in the body language, it's in the tone of voice. It may not be in the words themselves.* [P10]

## Qualitative Analysis of Nurse Interviews

The data from nurse interviews demonstrated that rapport develops when nurses know their patients as persons and is achievable in VCVs with some adaptations. Four key themes were identified during the analysis of the nurse data: (1) rapport building begins with nurses knowing their patients as persons, (2) much of the bedside manner can be translated into the webside manner, (3) differences in videoconferencing that may impact rapport are important to recognize, and (4) cultivating nurse-patient rapport in VCVs is essential for quality care and nurse job satisfaction. Multimedia Appendix 4 displays each theme, associated categories, codes, and additional participant quotations.

### *Theme 1: Rapport Building Begins With Nurses Knowing Their Patients as Persons*

This theme focuses on how nurses build rapport with their patients. Nurse participants described the importance of getting to know their patients holistically as individuals with needs, priorities, and lives apart from their cancer diagnosis. This connection facilitated trust and a sense of being well cared for that was valued by the nurse, the patient, and the patient's family.

#### Viewing the Patient From a Holistic Perspective

Oncology nurses must be knowledgeable about their patients' diagnoses and treatment plans and understand the impact of the disease on their quality of life. This holistic knowing was described by 1 participant:

> *We are here to talk about their health,…go through the review of systems and how they're feeling,…the labs,…the scan results, but I also like to say, "When you're not here, what are you doing outside of your cancer? Are you getting out of the house?" or "What's your family situation like? Are you working?" Just to know them on a more personal level and their families…the whole person, not just the cancer.* [Nurse 10 (N10)]

#### Attentiveness and Listening to Know the Patient and Understand Their Perspective

Nurses described how getting to know patients on a personal level creates a sense of bonding or rapport that facilitates understanding patient needs and allows the nurses to be empathetic and good patient advocates. One nurse participant explained:

> *Over the time that you spend with a patient in the clinic, you get to know them…their family, their likes, their dislikes, where they've traveled to. You make bonds. You might have similar situations or families. They tell you a lot of stories…I almost feel like I'm able to ask the physician a question on their behalf,…I try to be the patient advocate and put myself in the patient's shoes…It's like having a new friend that you just kind of learn everything about, and you want the best for them.* [N6]

#### Shared Experiences Enhance Trust, Open Communication, and Quality of Care

Nurses described intentionally asking patients questions about things they might have in common to facilitate a sense of connection. They often described rapport as foundational to building trust and open communication. One nurse noted:

> *Trust goes both ways. The patient has to trust the nurse to give them the correct care, and the nurse has to trust the patient to tell [them] when things are different or problematic. Some patients don't want to complain, and so they don't speak up when they have new pain or are more nauseous. If you have that rapport, you can kind of get to those underlying issues a little easier.* [N2]

Relationships built on rapport and trust were valued by nurses, in part because they believed this facilitates their ability to make good clinical decisions. One nurse stated:

> *When you connect with a patient, it's easier to navigate their care because you understand them. You understand if they're having certain symptoms, what it does and doesn't mean. Rapport really helps in terms of clinical judgment.* [N10]

### *Theme 2: Much of the Bedside Manner Can Be Translated Into the Webside Manner*

Theme 2 focuses on how nurses are adapting communication and relationship-building techniques to the videoconferencing setting, recently termed the "webside manner." Most of the nurses felt some extra effort is necessary to enhance connection and ensure humanization during VCVs. Many expressed confidence in their ability to create and maintain strong nurse-patient relationships in VCVs, especially with patients they knew from IPVs.

#### Similar Ability to Communicate as In-Person Visits

Nurses viewed their communication with patients in VCVs comparable to in-person communication. One nurse practitioner stated:

> *The rapport remained heavily intact throughout videoconferencing. I didn't feel I missed anything. Our conversations were very fruitful. We still had the same abilities to cover every topic.* [N5]

Some nurses even felt communication is enhanced as it allows them to schedule a VCV to share results or check on symptoms, which in the past might have been done by phone or email. One participant stated:

> *It actually improves our ability to communicate with patients. Oftentimes a patient I'm seeing virtually is not someone I would have brought into the office…so it allows a patient to communicate more with us than perhaps they would have otherwise.* [N3]

Another participant suggested that small adaptations to teaching are required for videoconferencing to be most effective:

> *If I'm doing a chemotherapy teach, I send them written instructions so that they can review them with me, and they can relay back. It helps if you communicate in multiple different ways at the same time.* [N8]

### Making Extra Efforts to Connect

Many strategies mentioned for creating a comfortable environment applied to VCVs and IPVs. One nurse explained:

> *I make more of an effort to have small talk over video. I need to bring everyone in, calm everyone down, and make this seem as comfortable as possible before we get to what we need to talk about.* [N2]

Some nurses related a need to take extra measures to connect with patients on videoconferencing to avoid a dehumanizing effect. One participant expressed:

> *If you don't come off human in a virtual box, then it just becomes like an automated survey type of appointment. "Where are you experiencing pain? What is the pain now? Okay, are you taking medications?"* [N12]

Nurses described asking patients about their home environment, family, and important personal events and using humor. One participant described helping patients feel at ease:

> *[Videoconferencing is] more awkward than real life, but…you can kind of have a laugh over [it]: "Oh, this is so awkward, but isn't nice that we didn't have to go out in the cold today?" Just stating the obvious is one way to kind of break the ice.* [N9]

All the nurses described the importance of using eye contact and body language to demonstrate attentiveness and compassion. As 1 participant described:

> *[By] maintaining eye contact, nodding, trying to use hand gestures, trying to limit distractions so that I can really focus, I can be more outward with my emotions and…convey compassion and empathy with my face.* [N3]

Some nurses described being extra alert and attentive during VCVs to compensate for cues that might be more obvious in an IPV.

### Ability to Form Relationships and Develop Rapport Remains Intact

Nurses reported an ability to build relationships and rapport over videoconferencing. One nurse stated:

> *We did the initial nursing assessment. We would talk about the emotional aspects, how patients were handling the diagnosis, how to start the treatments, and what that looked like. Overall, the rapport with the patients was not problematic.* [N1]

However, some nurses preferred IPVs for the initial visit and met patients in person whenever possible. One participant explained:

> *If it's a patient that I'm familiar with, it's a little easier [to establish rapport], but if it's a new patient that I've only met virtually during COVID, it's harder to build that rapport virtually, and if they're here to see the doctor and I can pop in, I try to pop in to meet them in person because I feel like it just adds to the rapport when you actually see them face to face.* [N4]

One benefit to video visits during COVID-19 was that nurses and patients could interact without wearing the face masks required for IPVs. Many nurses described the value of the unmasked interactions with patients afforded by VCVs. For example, a nurse noted:

> *I felt like we were able to develop that trusting relationship…Now that we are seeing people more regularly in person, some patients have actually said they appreciated the telehealth because I didn't have a mask on.* [N7]

## *Theme 3: Differences in Videoconferencing Visits That May Impact Rapport Are Important to Recognize*

This theme focuses on how nurses described some unique characteristics of videoconferencing and their personal experiences of how videoconferencing creates barriers to or facilitates care.

### There Is Only So Much You Can Do on Zoom

Nurses described clinical evaluations (eg, assessment techniques requiring hands-on physical examination) that were unfeasible in a VCV. As 1 nurse described:

> *There's no doubt about the fact that when you're in person, you know, you can put your hands on people. You can check an incision. You can listen to someone's lungs if they have a complaint. You can look in their mouth if they have a mouth sore. You can look at a rash.* [N9]

Participants considered some aspects of patient care, such as difficult or emotional conversations, best suited to IPVs. Nurses wanted to be with patients physically to evaluate their response and provide comfort and support. One nurse explained:

> *[What is lost in VCVs is] the body language and the cues that you would do in person. If you're having a difficult conversation, I would roll my chair over and bridge the gap physically…[During] a hospice talk, I might put a hand on a shoulder or on an arm or on a knee, offer tissues…show [in] nonverbal ways that I care, I'm concerned, and I'm really listening and engaged. I feel like you can only do so much on a Zoom [call].* [N4]

For many of the nurses, not being in the same physical space as the patient inhibited their ability to assess the encounter (eg, how the patient received information) thoroughly. The abrupt endings of video visits made some participants uncomfortable. One nurse participant explained:

> *Sometimes I leave a virtual visit not knowing how things landed with the patient because I'm not there, so if it's a big conversation, or a difficult conversation, or we're talking about a lot of things at once, sometimes it's harder to read the patient and how things are being processed. Whereas in person, before I leave a room, I usually make sure that things [are okay].* [N3]

### There's No Place Like Home

Nurses observed that patients seem more comfortable in their own homes during VCVs and appreciated the opportunity to see into the patients' home environment. As 1 nurse stated:

> *[Patients] may feel they have more time to answer questions when they're on the computer in their own environment and they're not stressed about coming into an appointment in an unfamiliar situation.* [N6]

### Barriers Related to Technology, Language, and Hearing

Nurses noted that the patient's level of experience with videoconferencing technology could present a barrier, and they had mixed impressions about its impact on rapport building. One nurse noted:

> *Patients have different levels of comfort and knowledge about the technology. Some…can't figure it out. Sometimes those barriers, like not being able to see them or make the eye contact, diminish rapport.* [N4]

Some nurses acknowledged difficulties but felt they do not interfere with rapport building:

> *If the Wi-Fi's bad or the connection's bad or…they can't hear me or I can't hear them, or if they're trying to bring up audio or the video. I think those can impact [the patient's stress level], but I'd say for the most part, I don't think I've had difficulty developing rapport with patients.* [N7]

Videoconferencing with patients who have hearing impairments or do not speak English presented an additional challenge according to some nurse participants. One nurse noted:

> *We have taken care of [some]ethnically diverse patients lately, and I find that's definitely more challenging on video because they're often not sitting in front of me. There could be a language barrier or they're just not quite understanding what you're trying to communicate. That definitely takes extra work, [and] you have to be sensitive to those things.* [N11]

## *Theme 4: Cultivating Nurse-Patient Rapport in Videoconferencing Is Essential for Quality Care and Nurse Job Satisfaction*

This theme focuses on practice adaptations to accommodate the addition of videoconferencing as a modality of care. Nurses expressed a desire for guidelines related to balancing patient convenience with the clinician's need to accomplish clinical goals. Maintaining a sense of connection or rapport with patients was described as important in ensuring quality care and the nurse's sense of fulfillment.

### Everyone Is Still Learning, but We Are Finding Ways to Adapt

Nurses acknowledged that using videoconferencing requires them to adapt, and some expressed mixed feelings. Overall, participants' descriptions suggested that they and their patients are acclimating. One participant said:

> *The pandemic taught us a lot…In health care at least, we learned there's a lot of things we can do, and how quickly we all acclimated is amazing, and so I just hope [VCVs] stay around.* [N5]

Another nurse described their experience of adapting to VCVs more tentatively:

> *It's kind of bumpy…I do prefer an in-person visit…but a telehealth visit with video is a lot better than just the telephone…It's still new, so I have mixed feelings about it.* [N6]

### Determining What Is Best for the Person and the Patient

Nurses recognized the advantages and limitations of videoconferencing, expressing that the best option differs from patient to patient and nurse to nurse, even for difficult conversations. Nurses commonly described a preference for sharing bad news with patients in person. One participant shared:

> *Someone asked me today, "Am I going to die?" and I moved closer to her…these moments of eye contact where you really have to make a connection with someone so that they know you're telling them the truth, that's hard over Zoom.* [N9]

However, 1 nurse described that patients and nurses are acclimating:

> *When we were at the beginning of the pandemic, we were worried that having a hospice conversation would be really hard over video…[but] if the patient's not feeling well, they don't want to come in. They have their family there,…the world is changing, and people are more and more used to having hard conversations over Zoom.* [N10]

Nurses reported a need to balance patient preferences with clinical evaluation, communication, and rapport factors. Some felt patients are the major beneficiaries of videoconferencing, with clinicians left to figure out how to adapt. Participants did view videoconferencing as an opportunity to follow patients more closely. Nurses felt that granting patients a choice over the format of visits is a means of empowering them.

The nurses interviewed hoped that VCVs would continue to be an option for seeing patients after the pandemic. As 1 nurse stated:

> *I hope...[for] sort of a hybrid where we can connect with our patients virtually, and…in person. We can be there in multiple ways, but always constantly there.* [N8]

### The Unique Challenges Presented by Videoconferencing Require Professional Development and Practice Guidelines

The collective experiences of the nurses in the study revealed a need and desire for practice guidelines for nurses and patients and professional development. As 1 nurse explained:

> *It is important to have guidelines…on an institutional level,…for patients to understand that it isn't the sole way that we can take care of them, . . . they do need to still see someone in person, because [otherwise] that could be dangerous. Things can be missed, so having those guidelines from an institutional level would…set the boundaries and the safety checkpoints to ensure that we're providing the best care we can for patients and in proper intervals.* [N5]

Several nurse participants expressed disappointment that as the pandemic waned, they were encouraged to bring patients to the clinic for IPVs when a VCV would have been sufficient. One nurse stated:

> *I think that [VCVs] serve a purpose for a very specific group of our patients…I'm not sure why it took the pandemic for us to realize that they were a really useful and appropriate way [to provide care]...They've been incredibly successful for providers and for patients, in certain settings, and I wish that we could use them more. I was kind of expecting a revolution to happen, and I'm really disappointed [this has not] played out.* [N9]

### Developing Rapport With Patients Is Important to Oncology Nurses

Without exception, the nurses interviewed described rapport as important to nursing practice and essential to high-quality care. Nurses described a sense of professional satisfaction from developing close relationships with their patients. As 1 nurse stated:

> *[Rapport] really can make my day. It's one of the best parts of the job, and that's probably what brings people to this line of work…It's a job with a lot of stress and it can be emotionally taxing…[H]aving that level of satisfaction and warmth and looking forward to seeing our patients really makes the job so enjoyable. It's very helpful in…grounding me in what I do and why I do it every day.* [N5]

### Comparison of Patient and Nurse Analyses

Comparison of the themes derived from the interviews showed similarities in how the nurses and patients described the experience of rapport during VCVs. Taken in totality, the data from all participants demonstrated that a person-centered and relationship-based approach can support nurse-patient rapport and the development of guidelines in videoconferencing for persons with cancer. Three themes fit the collective data: (1) person-centered and relationship-based care is valued and foundational to nurse-patient rapport in oncology ambulatory care regardless of how care is delivered, (2) adapting a bedside manner to facilitate rapport in VCVs is feasible, and (3) nurses and patients can work together to create options across the care trajectory that are person-centered and ensure quality care outcomes.

### *Theme 1: Person-Centered and Relationship-Based Care Is Valued and Foundational to Nurse-Patient Rapport in Oncology Ambulatory Care Regardless of How Care Is Delivered*

This theme focuses on the importance of person-centered and relationship-based care to both the nurse and the patient. The synergy between the patient's desire to be known on a personal level and the nurse's desire to know the patient provides a foundation for rapport building in oncology care delivered in an ambulatory care setting in both IPVs and VCVs. Patients and nurses used the words "treated as a person not just a patient" to describe a desire for holistic and personalized knowing. One patient emphasized the importance of the nurse and patient being authentic:

> *It's about me being a person and the nurse being a person rather than a job and a patient. You can't build rapport around a job and a patient. That's not doable, so anything that makes the nurse more of a person…She asks how things are going. I ask how things are going, and she answers…that little bit of back and forth makes a big difference…I trust a person, and I feel more seen as a person.* [P10]

Nurses and patients emphasized the importance of the nurse-patient relationship being built on mutual trust and understanding. This evolved out of knowing each other and the shared experience of managing a cancer diagnosis. One patient noted:

> *[Rapport is] having developed a relationship of trust and understanding…like you're on the same page. You understand each other…It's very important to have rapport because there are times when you really need somebody who understands what you're going through…so [you]'re not feeling alone and isolated.* [P1]

### *Theme 2: Adapting a Bedside Manner to Facilitate Rapport in Videoconferencing Visits Is Feasible*

This theme focuses on common strategies identified by patients and nurses to facilitate rapport and communication during VCVs. The behaviors nurses and patients described as important were the same for VCVs and IPVs; however, both nurses and patients seemed acutely aware of how body language in particular communicates attentiveness, compassion, and empathy in videoconferencing. One nurse described the importance of "maintaining eye contact, nodding, trying to use like hand gestures while I'm talking, trying to limit distractions in the visit so that I can really focus and…convey compassion and empathy with my face" [N3].

Most of the participants expressed that cultivating rapport and communicating in a VCV is feasible but that being able to have at least some IPVs is best for nurturing the nurse-patient relationship. As 1 nurse stated:

> *We see a lot of our patients frequently, and the ability to be able to alternate visits, like every other or every*

> *second visit in person and do the other ones on video has been a great benefit.* [N7]

### *Theme 3: Nurses and Patients Can Work Together to Create Options Across the Care Trajectory That Are Person-Centered and Ensure Quality Care Outcomes*

This theme highlights how both patients and nurses recognized that not all care can be provided in VCVs but that, when used selectively, good care can be provided in VCVs. As 1 nurse stated:

> *Patients [who] are established and they're just coming in for a scan review or a quick lab check, it just makes their life so much easier. You already have an established relationship, so it doesn't change anything.* [N10]

Similarly, a patient stated:

> *I find it's a really good way for communicating…[F]or certain visits, it really works.* [P10]

## Discussion

### Principal Findings

The purpose of this qualitative study was to describe the experiences of nurses and patients participating in oncology telehealth VCVs, with a focus on their ability to cultivate rapport. The findings suggest that rapport building is achievable within VCVs, with many traditional bedside, in-person strategies transferrable to the videoconferencing environment. There was a striking similarity in the descriptions of rapport building by patients and oncology nurses; both described a personal connection as foundational to building a trusting relationship and important for high-quality, satisfying care. Patients and nurses acknowledged that videoconferencing has benefits, challenges, and limitations but were interested in determining when and how to make optimal use of this new care modality.

Few studies to date have focused on ambulatory oncology nurse-patient rapport in VCVs. Consistent with previous research on relationship building in oncology IPVs [42-45], all participants in our study described the importance of the patient being known as a person rather than solely as a patient. Nurses described strategies to know the patient more holistically, and some patients appreciated knowing the nurse in a more personal way. Nurses described using self-disclosure, a common strategy for building relationships in nursing practice [64], to create a personal connection with their patients. Self-disclosure has been described as beneficial in building rapport with persons who have cancer [44,65].

Patients in our study valued being heard and appreciated the nurse taking time to provide information and answer their questions thoughtfully, and nurse participants described the importance of being attentive and listening closely. The importance of these behaviors is well described in the literature [39,44,48-50,66]. In a recent study exploring nursing listening behaviors, 70% of patients described nurse eye contact and attentiveness as an indication that the nurse was listening [66]. Conversing with patients on a personal level by asking and answering questions and providing a welcoming environment was described as a way to “transform an otherwise inauspicious moment into a powerful connection” [66]. Studies in oncology ambulatory care note that the absence of these caring behaviors leads patients to describe care as dehumanizing [67,68].

Our findings reflect attributes used to define rapport in the literature, including a shared experience comprising positive affect; mutual respect, acceptance, care, and concern; and behavioral synchrony [39,40,69]. Patient and nurse participants indicated that a shared connection facilitates outcomes similar to those described in the literature, including trust [46,66], open communication [39], comfort [66,70], confidence in the plan of care, and improved clinical judgment. Studies have suggested that rapport not only influences the patient’s perception of care but also has a tangible impact on care outcomes [41,71].

Both nurses and patients in our study felt that many of the strategies that build rapport in IPVs are effective in VCVs. Similarly, Elliott et al [72] suggested that patients who are satisfied with telemedicine encounters value their relational experience. The terms associated with their study’s code for “build rapport” align with the findings in our study, including “affective connection/comments of appreciation, trust-building, caring, concerned bedside manner, used nonverbal gestures that show care and concern, provided emotional support, understanding, developed a partnership, helpful, nice, friendly, easy to talk to” [72]. Other studies of videoconferencing have described the importance of knowing the patient as a person [73] and using eye contact, facial expressions [74,75], and other body language [76] in conveying attentiveness and emotions, including empathy. Our participants described the importance of nonverbal behaviors during VCVs to enhance communication and connection.

Although limited studies have explored relationship building in oncology VCVs, our study had findings similar to a qualitative study by van Gurp [24], which involved interviewing patients receiving care via videoconferencing with a palliative care team of whom over half were nurses. This study found that personalized patient-clinician relationships during videoconferencing are facilitated by consistent, empathetic engagement with the same clinician, and like our study, relationships were facilitated by clinicians who listened attentively and exchanged a mixture of medical and personal information during conversations.

Other nursing studies have described technology failures as a barrier to rapport [33,77]. Some participants in our study noted that disruptions in the home environment or internet connectivity have the potential to interfere with rapport, although this was not a significant concern. One patient participant expressed a preference to share confidential information with their nurse in person, due to uncertainty about videoconferencing security. In van Gurp’s [24] study, the ability to see but not touch made some palliative care clinicians reluctant to share difficult news (eg, a nurse expressed a desire to be able to provide physical comfort in these situations); however, for some patients, this physical separation made it easier for them to express feelings. These different perspectives were articulated by nurses and patients in our study. Some participants felt being in the comfort of their own homes made sharing emotions and concerns easier, while others felt it was important to have difficult news shared in person.

Although van Gurp's study [24] was the only research we identified that focused on rapport and videoconferencing in an oncology population, incidental findings from other nursing studies prepandemic suggest that setting up the environment to allow mutual attentiveness (eg, adjusting the camera position and audio volume) and ensuring privacy facilitate rapport in video visits [9,14,24,33,78-81]. Other studies show that a nurse's positive attitude toward technology [78] and the availability of others to assist patients onsite are helpful [33,78,82-84]. In our study, nurses and patients seemed equally willing to accept the challenges associated with videoconferencing, given the perceived benefits. There were relatively few concerns expressed about using the technology or its impact on relationship building. This phenomenon may be due to the increased need for and utilization of videoconferencing during the pandemic, especially for persons with cancer. In a study using videoconferencing with older adults experiencing depression, patients described the technical challenges as "little things," which the authors suggested was because the patients benefited from videoconferencing in many ways [85]. A central finding of this study was that an optimistic outlook on videoconferencing influenced the expectations and attitudes of both patient and clinician participants, mitigating negative feelings about technological challenges [85].

Concerns that videoconferencing would depersonalize care [19-22] were not supported by our findings. Although patients and nurses expressed that IPVs are more desirable in certain situations, none of the participants described their VCVs as impersonal; on the contrary, many described their ability to develop rapport and communicate effectively during VCVs. According to Barrett's [33] ground theory, the primary function of the nurse in videoconferencing is to provide an operational, clinical, therapeutic, and social presence. The nurse cultivates rapport in the nurse-patient relationship during videoconferencing by providing reassurance and support (therapeutic presence) and creating connection, thus "giving the patient the sense they are 'in the room' with them" (social presence) [33]. Although the word "presence" was not used to describe behavior or attributes by any of our participants, the need to have more focused attention was described by several of the nurses and 1 patient. The importance of telepresence, generally described as feeling physically present during a computer-mediated encounter [86], has been the focus of some research [21,33,36].

Our findings suggest that oncology nurses and persons with cancer are receptive to integrating videoconferencing into the care trajectory but feel it is important to determine what type of visit will best serve the patient for each care encounter. This like-mindedness bodes well for nurse-patient collaboration on decisions about whether a visit should occur virtually or in person at the cancer center. Similar to our results, recent studies indicate that both patients and clinicians are increasingly receptive to oncology care provision through VCVs and want the option of videoconferencing to extend post–COVID-19 [87,88]. As in previous nursing studies, both patients and nurses in our study acknowledged that video visits could not replace all in-person encounters [24,33,79,83,85,89]. This was felt to be especially true for serious conversations [14,24] and initial encounters [83,85]. At least 1 nurse in the study looked for opportunities to physically see patients during their chemotherapy treatments whom they had only met in VCVs in order to enhance their connection. Although the study did not ask patients during the interview about how long they had known their nurses, 50% of the patient participants had been in treatment for less than a year. Most, though not all, patients and nurses in the study, felt IPVs visits are better for more difficult conversations. Our findings and those of previous studies suggest that decisions about whether to use videoconferencing must consider the patient's preferences and everchanging needs. This conclusion is supported by the Institute for Healthcare Improvement's (IHI) recent white paper on telemedicine [90], leading some to suggest that decisions regarding virtual care require the following amendment to the precision medicine maxim: "Provide the right treatment, to the right patient, at the right time, and in the right place" [91].

## Implications for Practice, Policy, and Professional Development

We recommend that policy and practice guidelines be person centered, allowing clinicians to assess in real time the type of visit that can best meet the patient's holistic needs. The IHI has proposed a framework for telemedicine that is safe, equitable, and person centered: "Honoring the patient's wishes as long as those desires are consistent with delivering safe and effective care" [90]. Our results suggest that nurses and persons with cancer have an appreciation for the challenges as well as the benefits of VCVs and understand that this environment may not always provide the best option for high-quality, safe care.

Although the pandemic forced a rapid adoption of video visits, clinicians remain uncertain about how best to provide virtual care. In a recent mixed methods study by Elsevier Health of 3000 nurses and doctors, over half of the clinicians felt telehealth would negatively impact their ability to demonstrate empathy and requested guidance on learning webside skills [51]. Although most of our participants felt able to transfer many bedside skills into VCVs, there is mixed evidence in the literature on whether videoconferencing can produce the same empathetic experience as an IPV [92]. Notable efforts have been made to provide guidance on a webside manner [26-28,93,94], but additional research is needed to ensure that professional development and practice guidelines are evidence based, and our findings provide data regarding relationship-based care and patient-centered communication in a videoconferencing environment. Schools of nursing and medicine need to prepare future practitioners to care for patients in a virtual health care environment, with competencies specific to these digital tools [92,95]. Fortunately, nursing theories [22,33,35] and conceptual models [96] exist to support this important work, and relationship-based care has been successfully incorporated into health care clinician curriculum and professional practice models [97,98], showing improvement in health care delivery [41,99,100].

Our findings identified barriers to rapport building in videoconferencing that have been previously identified in studies on videoconferencing, including interruptions due to breaks in the internet connection [32,33,77,85,90], concerns about privacy, [9,12,24,83], and the limitations imposed by a lack of physical presence [24,32,33,83,90]. Nurses and patients in our study recognized that balancing these challenges and limitations with the benefits of videoconferencing is an essential competency requiring additional research and guidelines.

### Future Research

Research is needed to better understand how specific rapport-building strategies can be translated or adapted to VCVs, including nonverbal communication and active listening techniques. Additionally, exploring how various types of visits (eg, first encounters, delivery of bad news) and contextual factors (eg, virtual backgrounds, quality of connectivity, sensory and language barriers) influence rapport building in videoconferencing would be useful.

Ensuring adequate access to VCVs is multifaceted and requires not only that adequate devices and Wi-Fi resources be available and affordable but also that patients and providers have the skills and support to incorporate them into their care services [87,101]. Investigating how digital literacy and access to telehealth technology influence patient and nurse utilization of VCVs, while not addressed in this study, must be a focus of future studies if videoconferencing is to become a mainstay of health care. Existing health and technological disparities became more apparent during COVID-19 [101,102], and future studies are needed to better understand the relationship between the services offered and the needs and abilities of patients and providers. For example, the lack of access to adequate devices and Wi-Fi resources for VCVs is well documented [1,103,104], but the impact of other factors, such as language barriers and availability of privacy and safe spaces for VCVs [105], require better understanding and accommodation. Specifically in relationship to rapport, there is some evidence that qualitative differences between type of device, strength of broadband, and level of literacy impact the level and quality of empathetic communication within VCVs [92]. Addressing these factors is essential if we are to avoid increasing the digital divide among populations who may already be at a higher risk for cancer due to social and economic disparities [106,107]. Infusing a health equity lens as we generate new knowledge about VCVs can help prevent the augmentation of health disparities and promote health and scientific equity, especially for groups that are underrepresented.

Other interesting questions generated from our findings include the role of nurse self-disclosure on the development of rapport and whether videoconferencing can create a safer space or reduce the power differential between clinicians and patients. Our study shows the importance of rapport in the nurse-patient relationship and its impact on patient care and nursing job satisfaction, highlighting the need for additional research in this area.

### Strengths and Limitations

The strengths of this study include the collection of data from persons with cancer and oncology nurses. Multiple strategies were used to improve the trustworthiness of findings. The analysis included categories covering a wide range of participants' responses to achieve credibility. Dependability was ensured by selecting quotations from multiple participants and identifying how these were linked to results. To allow the findings to be transferred or applied to other settings or groups, the context of the study and participant characteristics were thoroughly described.

This study has several limitations. First, the data analysis and its interpretation depended on the researchers' skills, assumptions, and experience. Second, data were collected from 1 cancer center and were therefore influenced by the organizational system and its practices. Third, to minimize the burden on the participants, member checking was not used after the data were analyzed; however, the researcher conducting the interviews frequently validated their understanding of participants' answers during the interview. Fourth, despite our efforts to include nurses and patients who might otherwise have been underrepresented in the study, the study participants mostly self-identified as female, were White/Caucasian, and were employed full-time, with at least some college education. There was a wide range of experience with VCVs among the participants; however, most were using computer technology and videoconferencing for work and personal affairs, suggesting their technological skills, devices, and internet access might not be representative of most patients with cancer and their nurses. This limits the ability to generalize the study findings to patients and providers with less access to and experience with VCV technology.

### Conclusion

Although providing care within the videoconferencing environment may require adapting practices, the essential nature of nursing need not be affected. In this study, the overall synergy between the nurse and patient data and the specific descriptions from patients and nurses on ways to establish rapport were striking. Patients and nurses considered rapport essential to the nurse-patient relationship and high-quality care, thus affirming nursing's commitment to person-centered care and the profession's capacity to understand patient needs holistically. Nurses are well positioned to assume an advocacy role for care across the cancer continuum and leadership in the development of guidelines, policies, and future research inquiry on VCVs. Contrary to concerns that videoconferencing would be impersonal and inhibit rapport and relationship building, this study indicated that rapport can be established during VCVs and that many of the strategies used during IPVs are equally successful in videoconferencing. Findings from this study provide some of the descriptive research necessary for the development of evidence-based practice guidelines and interventions to support nurse-patient therapeutic relationships during VCVs.

## Authors' Contributions

PDK, JCD, and TJ designed the study protocol. PDK interviewed all study participants. PDK and JCD coded and analyzed the data. PDK drafted the initial manuscript, and all other authors contributed to critical reviews and revisions of the manuscript. All authors have approved the final manuscript and agreed to publication.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of patient participants.
[DOCX File , 16 KB - jmir_v24i9e39920_app1.docx ]

Multimedia Appendix 2
Characteristics of oncology nurse participants.
[DOCX File , 16 KB - jmir_v24i9e39920_app2.docx ]

Multimedia Appendix 3
Patient data analysis.
[DOCX File , 24 KB - jmir_v24i9e39920_app3.docx ]

Multimedia Appendix 4
Nurse data analysis.
[DOCX File , 27 KB - jmir_v24i9e39920_app4.docx ]

## References

1. Knudsen KE, Willman C, Winn R. Optimizing the use of telemedicine in oncology care: postpandemic opportunities. Clin Cancer Res 2021 Feb 15;27(4):933-936 [FREE Full text] [doi: 10.1158/1078-0432.CCR-20-3758] [Medline: 33229457]
2. Alpert JM, Taylor G, Hampton CN, Paige S, Markham MJ, Bylund CL. Clinicians' perceptions of the benefits and challenges of teleoncology as experienced through the COVID-19 pandemic: qualitative study. JMIR Cancer 2022 Feb 24;8(1):e34895 [FREE Full text] [doi: 10.2196/34895] [Medline: 35142622]
3. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
4. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: a review of reviews and applications in line with guidelines and recommendations. Digit Health 2019;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
5. Penny RA, Bradford NK, Langbecker D. Registered nurse and midwife experiences of using videoconferencing in practice: a systematic review of qualitative studies. J Clin Nurs 2018 Mar 11;27(5-6):e739-e752. [doi: 10.1111/jocn.14175] [Medline: 29149507]
6. Speyer R, Denman D, Wilkes-Gillan S, Chen Y, Bogaardt H, Kim J, et al. Effects of telehealth by allied health professionals and nurses in rural and remote areas: a systematic review and meta-analysis. J Rehabil Med 2018 Feb 28;50(3):225-235 [FREE Full text] [doi: 10.2340/16501977-2297] [Medline: 29257195]
7. Kitamura C, Zurawel-Balaura L, Wong RK. How effective is video consultation in clinical oncology? A systematic review. Curr Oncol 2010 Jun 26;17(3):17-27 [FREE Full text] [doi: 10.3747/co.v17i3.513] [Medline: 20567623]
8. McGrowder DA, Miller FG, Vaz K, Anderson Cross M, Anderson-Jackson L, Bryan S, et al. The utilization and benefits of telehealth services by health care professionals managing breast cancer patients during the COVID-19 pandemic. Healthcare (Basel) 2021 Oct 19;9(10):1401 [FREE Full text] [doi: 10.3390/healthcare9101401] [Medline: 34683081]
9. Funderskov KF, Boe Danbjørg D, Jess M, Munk L, Olsen Zwisler A, Dieperink KB. Telemedicine in specialised palliative care: healthcare professionals' and their perspectives on video consultations-A qualitative study. J Clin Nurs 2019 Nov 09;28(21-22):3966-3976. [doi: 10.1111/jocn.15004] [Medline: 31328336]
10. Jess M, Timm H, Dieperink KB. Video consultations in palliative care: a systematic integrative review. Palliat Med 2019 Sep 12;33(8):942-958. [doi: 10.1177/0269216319854938] [Medline: 31187688]
11. Steindal SA, Nes AAG, Godskesen TE, Dihle A, Lind S, Winger A, et al. Patients' experiences of telehealth in palliative home care: scoping review. J Med Internet Res 2020 May 05;22(5):e16218 [FREE Full text] [doi: 10.2196/16218] [Medline: 32369037]

12. Funderskov KF, Raunkiær M, Danbjørg DB, Zwisler A, Munk L, Jess M, et al. Experiences with video consultations in specialized palliative home-care: qualitative study of patient and relative perspectives. J Med Internet Res 2019 Mar 21;21(3):e10208 [FREE Full text] [doi: 10.2196/10208] [Medline: 30896436]
13. Bradford N, Young J, Armfield NR, Bensink ME, Pedersen L, Herbert A, et al. A pilot study of the effectiveness of home teleconsultations in paediatric palliative care. J Telemed Telecare 2012 Dec;18(8):438-442. [doi: 10.1258/jtt.2012.gth103] [Medline: 23155114]
14. Oliver D, Washington K, Wittenberg-Lyles E, Demiris G, Porock D. 'They're part of the team': participant evaluation of the ACTIVE intervention. Palliat Med 2009 Sep 14;23(6):549-555 [FREE Full text] [doi: 10.1177/0269216309105725] [Medline: 19443524]
15. Melton L, Brewer B, Kolva E, Joshi T, Bunch M. Increasing access to care for young adults with cancer: results of a quality-improvement project using a novel telemedicine approach to supportive group psychotherapy. Pall Supp Care 2016 Jul 26;15(2):176-180 [FREE Full text] [doi: 10.1017/s1478951516000572]
16. Nordtug B, Brataas HV, Rygg LO. Patient experiences with videoconferencing as social contact and in follow-up from oncology nurses in primary health care. Health Psychol Open 2021 Apr 23;8(1):20551029211012208 [FREE Full text] [doi: 10.1177/20551029211012208] [Medline: 33996137]
17. McCabe HM, Smrke A, Cowie F, White J, Chong P, Lo S, et al. What matters to us: impact of telemedicine during the pandemic in the care of patients with sarcoma across Scotland. JCO Global Oncol 2021 Dec(7):1067-1073. [doi: 10.1200/go.20.00599]
18. Calton B, Shibley WP, Cohen E, Pantilat SZ, Rabow MW, O'Riordan DL, et al. Patient and caregiver experience with outpatient palliative care telemedicine visits. Palliat Med Rep 2020 Dec 01;1(1):339-346 [FREE Full text] [doi: 10.1089/pmr.2020.0075] [Medline: 34223495]
19. Brahnam S. Comparison of in-person and screen-based analysis using communication models: a first step toward the psychoanalysis of telecommunications and its noise. Psychoanal Perspect 2017 Apr 25;14(2):138-158. [doi: 10.1080/1551806x.2017.1304112]
20. Chhabra AM, Chowdhary M, Choi JI, Hasan S, Press RH, Simone CB. A national survey of radiation oncology experiences completing tele-consultations during the coronavirus disease (COVID-19) pandemic. Adv Radiat Oncol 2021 Jan;6(1):100611 [FREE Full text] [doi: 10.1016/j.adro.2020.10.024] [Medline: 33458451]
21. Grumme V, Barry C, Gordon S, Ray M. On virtual presence. Adv Nurs Sci 2016;39(1):48-59. [doi: 10.1097/ans.0000000000000103]
22. Locsin RC, Purnell M. Advancing the theory of technological competency as caring in nursing: the universal technological domain. Int J Hum Caring 2015 Jan;19(2):50-54. [doi: 10.20467/1091-5710-19.2.50]
23. Lally K, Kematick BS, Gorman D, Tulsky J. Rapid conversion of a palliative care outpatient clinic to telehealth. JCO Oncol Pract 2021 Jan;17(1):e62-e67. [doi: 10.1200/op.20.00557]
24. van Gurp J, van Selm M, Vissers K, van Leeuwen E, Hasselaar J. How outpatient palliative care teleconsultation facilitates empathic patient-professional relationships: a qualitative study. PLoS One 2015 Apr 22;10(4):e0124387 [FREE Full text] [doi: 10.1371/journal.pone.0124387] [Medline: 25902263]
25. Car J, Koh GC, Foong PS, Wang CJ. Video consultations in primary and specialist care during the covid-19 pandemic and beyond. BMJ 2020 Oct 20;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
26. Chua IS, Jackson V, Kamdar M. Webside manner during the COVID-19 pandemic: maintaining human connection during virtual visits. J Palliat Med 2020 Nov 01;23(11):1507-1509. [doi: 10.1089/jpm.2020.0298] [Medline: 32525744]
27. Webb M, Hurley SL, Gentry J, Brown M, Ayoub C. Best practices for using telehealth in hospice and palliative care. J Hosp Palliat Nurs 2021 Apr 8;23(3):277-285. [doi: 10.1097/njh.0000000000000753]
28. Banerjee SC, Staley JM, Howell F, Malling C, Moreno A, Kotsen C, et al. Communicating effectively via tele-oncology (Comskil TeleOnc): a guide for best practices for communication skills in virtual cancer care. J Cancer Educ 2021 Feb 05:1-6 [FREE Full text] [doi: 10.1007/s13187-021-01959-7] [Medline: 33544315]
29. Dalley D, Rahman R, Ivaldi A. Health care professionals' and patients' management of the interactional practices in telemedicine videoconferencing: a conversation analytic and discursive systematic review. Qual Health Res 2021 Mar 01;31(4):804-814. [doi: 10.1177/1049732320942346] [Medline: 32741261]
30. Henry B, Ames L, Block D, Vozenilek J. Experienced practitioners' views on interpersonal skills in telehealth delivery. Internet J Allied Health Sci Pract 2018;16(2):Art 2. [doi: 10.46743/1540-580x/2018.1745]
31. Newcomb AB, Duval M, Bachman SL, Mohess D, Dort J, Kapadia MR. Building rapport and earning the surgical patient's trust in the era of social distancing: teaching patient-centered communication during video conference encounters to medical students. J Surg Educ 2021 Jan;78(1):336-341. [doi: 10.1016/j.jsurg.2020.06.018]
32. Shaw SE, Seuren LM, Wherton J, Cameron D, A'Court C, Vijayaraghavan S, et al. Video consultations between patients and clinicians in diabetes, cancer, and heart failure services: linguistic ethnographic study of video-mediated interaction. J Med Internet Res 2020 May 11;22(5):e18378 [FREE Full text] [doi: 10.2196/18378] [Medline: 32391799]
33. Barrett D. Rethinking presence: a grounded theory of nurses and teleconsultation. J Clin Nurs 2017 Oct 24;26(19-20):3088-3098. [doi: 10.1111/jocn.13656] [Medline: 27874982]

34. Tuxbury J. The experience of presence among telehealth nurses. J Nurs Res 2013 Sep;21(3):155-161. [doi: 10.1097/jnr.0b013e3182a0b028] [Medline: 23958604]
35. Groom LL, Brody AA, Squires AP. Defining telepresence as experienced in telehealth encounters: a dimensional analysis. J Nurs Scholarsh 2021 Nov 31;53(6):709-717. [doi: 10.1111/jnu.12684] [Medline: 34060218]
36. Sitzman, K. Evolution of Watson's human caring science in the digital age. Int J Hum Caring 2017 Feb 01;21(1):46-52. [doi: 10.20467/1091-5710.21.1.46]
37. Liu X, Sawada Y, Takizawa T, Sato H, Sato M, Sakamoto H, et al. Doctor-patient communication: a comparison between telemedicine consultation and face-to-face consultation. Intern Med 2007;46(5):227-232 [FREE Full text] [doi: 10.2169/internalmedicine.46.1813] [Medline: 17329917]
38. Heckemann B, Wolf A, Ali L, Sonntag SM, Ekman I. Discovering untapped relationship potential with patients in telehealth: a qualitative interview study. BMJ Open 2016 Mar 02;6(3):e009750 [FREE Full text] [doi: 10.1136/bmjopen-2015-009750] [Medline: 26936904]
39. Epstein R, Street R. Patient-Centered Communication in Cancer Care: Promoting Healing and Reducing Suffering. Bethesda, MD: National Cancer Institute; 2007:1-203.
40. Tickle-Degnen L, Rosenthal R. The nature of rapport and its nonverbal correlates. Psychol Inquiry 1990 Oct;1(4):285-293. [doi: 10.1207/s15327965pli0104_1]
41. Haverfield MC, Tierney A, Schwartz R, Bass MB, Brown-Johnson C, Zionts DL, et al. Can patient-provider interpersonal interventions achieve the quadruple aim of healthcare? A systematic review. J Gen Intern Med 2020 Jul 09;35(7):2107-2117 [FREE Full text] [doi: 10.1007/s11606-019-05525-2] [Medline: 31919725]
42. Aldaz BE, Treharne GJ, Knight RG, Conner TS, Perez D. Oncology healthcare professionals' perspectives on the psychosocial support needs of cancer patients during oncology treatment. J Health Psychol 2017 Sep 01;22(10):1332-1344. [doi: 10.1177/1359105315626999] [Medline: 26837692]
43. Boston P, Bruce A. Palliative care nursing, technology, and therapeutic presence: are they reconcilable? J Palliat Care 2018 Dec 19;30(4):291-293. [doi: 10.1177/082585971403000409]
44. Thorne SE, Kuo M, Armstrong E, McPherson G, Harris SR, Hislop TG. 'Being known': patients' perspectives of the dynamics of human connection in cancer care. Psychooncology 2005 Oct;14(10):887-898; discussion 899. [doi: 10.1002/pon.945] [Medline: 16200520]
45. Grover C, Mackasey E, Cook E, Nurse H, Tremblay L, Clinician N, et al. Patient-reported care domains that enhance the experience of "being known" in an ambulatory cancer care centre. Can Oncol Nurs J 2018 Jul 19;28(3):166-171 [FREE Full text] [doi: 10.5737/23688076283166171] [Medline: 31148824]
46. O'Grady C, Dahm MR, Roger P, Yates L. Trust, talk and the dictaphone: tracing the discursive accomplishment of trust in a surgical consultation. Discourse Soc 2013 Aug 08;25(1):65-83. [doi: 10.1177/0957926513496354]
47. Crawford T, Roger P, Candlin S. Tracing the discursive development of rapport in intercultural nurse-patient interactions. Int J Appl Linguist 2017 Feb 07;27(3):636-650. [doi: 10.1111/ijal.12166]
48. Komatsu H, Yagasaki K. The power of nursing: guiding patients through a journey of uncertainty. Eur J Oncol Nurs 2014 Aug;18(4):419-424 [FREE Full text] [doi: 10.1016/j.ejon.2014.03.006] [Medline: 24721178]
49. Radwin LE, Farquhar SL, Knowles MN, Virchick BG. Cancer patients' descriptions of their nursing care. J Adv Nurs 2005 Apr;50(2):162-169. [doi: 10.1111/j.1365-2648.2005.03375.x] [Medline: 15788080]
50. Rchaidia L, Dierckx de Casterlé B, De Blaeser L, Gastmans C. Cancer patients' perceptions of the good nurse: a literature review. Nurs Ethics 2009 Sep 11;16(5):528-542. [doi: 10.1177/0969733009106647] [Medline: 19671641]
51. Goodchild L, Mulligan A, Shearing GE, Mueller T. Clinician of the Future; a 2022 Report. 2022. URL: https://www.elsevier.com/connect/clinician-of-the-future [accessed 2022-08-16]
52. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Sep;69(686):e595-e604 [FREE Full text] [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
53. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g]
54. Sandelowski M. What's in a name? Qualitative description revisited. Res Nurs Health 2010 Feb;33(1):77-84. [doi: 10.1002/nur.20362] [Medline: 20014004]
55. Koppel PD, De Gagne JC. Exploring nurse and patient experiences of developing rapport during oncology ambulatory care videoconferencing visits: protocol for a qualitative study. JMIR Res Protoc 2021 Jun 14;10(6):e27940 [FREE Full text] [doi: 10.2196/27940] [Medline: 34125073]
56. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
57. Sandelowski M. Sample size in qualitative research. Res Nurs Health 1995 Apr;18(2):179-183 [FREE Full text] [doi: 10.1002/nur.4770180211] [Medline: 7899572]
58. Marshall C, Rossman GB. Designing Qualitative Research. Sixth edition. Los Angeles: SAGE Publications; 2016:1-352.
59. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]

60. Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. Int J Qual Methods 2016 Nov 30;1(2):13-22. [doi: 10.1177/160940690200100202]
61. Elo S, Kääriäinen M, Kanste O, Pölkki T, Utriainen K, Kyngäs H. Qualitative content analysis: a focus on trustworthiness. SAGE Open 2014 Feb 11;4(1):215824401452263. [doi: 10.1177/2158244014522633]
62. Tracy SJ. Qualitative quality: eight "big-tent" criteria for excellent qualitative research. Qual Inq 2010 Oct 01;16(10):837-851. [doi: 10.1177/1077800410383121]
63. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
64. Unhjem JV, Vatne S, Hem MH. Transforming nurse-patient relationships-a qualitative study of nurse self-disclosure in mental health care. J Clin Nurs 2018 Mar 08;27(5-6):e798-e807. [doi: 10.1111/jocn.14191] [Medline: 29193417]
65. Radwin L. Oncology patients' perceptions of quality nursing care. Res Nurs Health 2000 Jun;23(3):179-190. [doi: 10.1002/1098-240x(200006)23:3<179::aid-nur2>3.0.co;2-t]
66. Loos N. Nurse listening as perceived by patients: how to improve the patient experience, keep patients safe, and raise HCAHPS scores. J Nurs Adm 2021;51(6):324-328. [doi: 10.1097/nna.0000000000001021]
67. Prip A, Pii KH, Møller KA, Nielsen DL, Thorne SE, Jarden M. Observations of the communication practices between nurses and patients in an oncology outpatient clinic. Eur J Oncol Nurs 2019 Jun;40:120-125. [doi: 10.1016/j.ejon.2019.03.004] [Medline: 31229201]
68. McIlfatrick S, Sullivan K, McKenna H, Parahoo K. Patients' experiences of having chemotherapy in a day hospital setting. J Adv Nurs 2007 Aug;59(3):264-273. [doi: 10.1111/j.1365-2648.2007.04324.x] [Medline: 17590208]
69. Fredrickson B. Positivity resonance as a fresh, evidence-based perspective on an age-old topic. In: Barrett LF, Haviland JM, editors. Handbook of Emotions. 4th ed. New York, NY: Guilford Press; 2016:847-858.
70. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
71. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014 Apr 9;9(4):e94207 [FREE Full text] [doi: 10.1371/journal.pone.0094207] [Medline: 24718585]
72. Elliott T, Tong I, Sheridan A, Lown BA. Beyond convenience: patients' perceptions of physician interactional skills and compassion via telemedicine. Mayo Clin Proc Innov Qual Outcomes 2020 Jun;4(3):305-314 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.04.009] [Medline: 32542222]
73. Rose S, Hurwitz HM, Mercer MB, Hizlan S, Gali K, Yu P, et al. Patient experience in virtual visits hinges on technology and the patient-clinician relationship: a large survey study with open-ended questions. J Med Internet Res 2021 Jun 21;23(6):e18488 [FREE Full text] [doi: 10.2196/18488] [Medline: 34152276]
74. Schmidt KL, Gentry A, Monin JK, Courtney KL. Demonstration of facial communication of emotion through telehospice videophone contact. Telemed J E Health 2011 Jun;17(5):399-401 [FREE Full text] [doi: 10.1089/tmj.2010.0190] [Medline: 21492032]
75. Bohannon LS, Herbert AM, Pelz JB, Rantanen EM. Eye contact and video-mediated communication: a review. Displays 2013 Apr;34(2):177-185. [doi: 10.1016/j.displa.2012.10.009]
76. Gupta A, Harris S, Naina HV. The impact of physician posture during oncology patient encounters. J Cancer Educ 2015 Jun 12;30(2):395-397. [doi: 10.1007/s13187-015-0807-2] [Medline: 25757904]
77. Hibbert D, Mair FS, May CR, Boland A, O'Connor J, Capewell S, et al. Health professionals' responses to the introduction of a home telehealth service. J Telemed Telecare 2004 Jun 24;10(4):226-230. [doi: 10.1258/1357633041424386] [Medline: 15273033]
78. Sävenstedt S, Zingmark K, Hydén LC, Brulin C. Establishing joint attention in remote talks with the elderly about health: a study of nurses' conversation with elderly persons in teleconsultations. Scand J Caring Sci 2005 Dec;19(4):317-324. [doi: 10.1111/j.1471-6712.2005.00346.x] [Medline: 16324054]
79. Hastings SN, Mahanna EP, Berkowitz TSZ, Smith VA, Choate AL, Hughes JM, et al. Video-enhanced care management for medically complex older adults with cognitive impairment. J Am Geriatr Soc 2021 Jan 23;69(1):77-84 [FREE Full text] [doi: 10.1111/jgs.16819] [Medline: 32966603]
80. Sharma U, Clarke M. Nurses' and community support workers' experience of telehealth: a longitudinal case study. BMC Health Serv Res 2014 Apr 10;14:164 [FREE Full text] [doi: 10.1186/1472-6963-14-164] [Medline: 24721599]
81. Oliver DP, Albright DL, Kruse RL, Wittenberg-Lyles E, Washington K, Demiris G. Caregiver evaluation of the ACTIVE intervention: "it was like we were sitting at the table with everyone". Am J Hosp Palliat Care 2014 Jun 26;31(4):444-453 [FREE Full text] [doi: 10.1177/1049909113490823] [Medline: 23713130]
82. Foster PH, Whitworth JM. The role of nurses in telemedicine and child abuse. Comput Inform Nurs 2005;23(3):127-131. [doi: 10.1097/00024665-200505000-00007] [Medline: 15900169]

83. Read Paul L, Salmon C, Sinnarajah A, Spice R. Web-based videoconferencing for rural palliative care consultation with elderly patients at home. Support Care Cancer 2019 Sep 7;27(9):3321-3330. [doi: 10.1007/s00520-018-4580-8] [Medline: 30613908]
84. Tachakra S, Rajani R. Social presence in telemedicine. J Telemed Telecare 2002 Jun 24;8(4):226-230. [doi: 10.1258/135763302320272202] [Medline: 12217106]
85. Christensen LF, Wilson R, Hansen JP, Nielsen CT, Gildberg FA. A qualitative study of patients' and providers' experiences with the use of videoconferences by older adults with depression. Int J Ment Health Nurs 2021 Apr 11;30(2):427-439. [doi: 10.1111/inm.12803] [Medline: 33179416]
86. Draper JV, Kaber DB, Usher JM. Telepresence. Hum Factors 1998 Sep 23;40(3):354-375. [doi: 10.1518/001872098779591386] [Medline: 9849099]
87. Turner K, Bobonis Babilonia M, Naso C, Nguyen O, Gonzalez BD, Oswald LB, et al. Health care providers' and professionals' experiences with telehealth oncology implementation during the COVID-19 pandemic: a qualitative study. J Med Internet Res 2022 Jan 19;24(1):e29635 [FREE Full text] [doi: 10.2196/29635] [Medline: 34907900]
88. Darcourt JG, Aparicio K, Dorsey PM, Ensor JE, Zsigmond EM, Wong ST, et al. Analysis of the implementation of telehealth visits for care of patients with cancer in Houston during the COVID-19 pandemic. JCO Oncol Pract 2021 Jan;17(1):e36-e43 [FREE Full text] [doi: 10.1200/OP.20.00572] [Medline: 33026951]
89. Lindberg B, Axelsson K, Öhrling K. Taking care of their baby at home but with nursing staff as support: the use of videoconferencing in providing neonatal support to parents of preterm infants. J Neonatal Nurs 2009 Apr;15(2):47-55. [doi: 10.1016/j.jnn.2009.01.004]
90. Perry A, Federico F, Huebner J. Telemedicine: Ensuring Safe, Equitable, Person-Centered Virtual Care. 2021. URL: http://www.ihi.org/resources/Pages/IHIWhitePapers/telemedicine-safe-equitable-person-centered-virtual-care.aspx [accessed 2022-08-16]
91. Binder AF, Handley NR, Wilde L, Palmisiano N, Lopez AM. Treating hematologic malignancies during a pandemic: utilizing telehealth and digital technology to optimize care. Front Oncol 2020 Jun 26;10:1183 [FREE Full text] [doi: 10.3389/fonc.2020.01183] [Medline: 32676459]
92. Myronuk L. Effect of telemedicine via videoconference on provider fatigue and empathy: implications for the Quadruple Aim. Healthc Manage Forum 2022 May 15;35(3):174-178. [doi: 10.1177/08404704211059944] [Medline: 35289218]
93. Chike-Harris KE, Garber K, Derouin A. Telehealth educational resources for graduate nurse faculty. Nurse Educ 2021 Jul 8;46(5):295-299. [doi: 10.1097/nne.0000000000001055]
94. Allen Watts K, Malone E, Dionne-Odom JN, McCammon S, Currie E, Hicks J, et al. Can you hear me now?: improving palliative care access through telehealth. Res Nurs Health 2021 Feb;44(1):226-237 [FREE Full text] [doi: 10.1002/nur.22105] [Medline: 33393704]
95. Cheng C, Humphreys H, Kane B. Transition to telehealth: engaging medical students in telemedicine healthcare delivery. Ir J Med Sci 2021 Oct 09:1-18 [FREE Full text] [doi: 10.1007/s11845-021-02720-1] [Medline: 34626350]
96. Nagel DA, Penner JL. Conceptualizing telehealth in nursing practice: advancing a conceptual model to fill a virtual gap. J Holist Nurs 2016 Mar 09;34(1):91-104. [doi: 10.1177/0898010115580236] [Medline: 25858897]
97. Haertl KL, Theismann A. Perceived impact of a relationship-based care curriculum. Creat Nurs 2022 Feb 16;28(1):36-41. [doi: 10.1891/cn-2021-0010]
98. Reilly J, Krause K, Vande Zande C, Knutzen B. Implementing relationship-based care as a professional practice model: promoting nurses' understanding and confidence to apply in practice. Creat Nurs 2019 Nov 01;25(4):e36-e43. [doi: 10.1891/1078-4535.25.4.e36] [Medline: 31796625]
99. Zulman DM, Haverfield MC, Shaw JG, Brown-Johnson CG, Schwartz R, Tierney AA, et al. Practices to foster physician presence and connection with patients in the clinical encounter. JAMA 2020 Jan 07;323(1):70-81. [doi: 10.1001/jama.2019.19003] [Medline: 31910284]
100. Ansari N, Wilson CM, Heneghan MB, Supiano K, Mooney K. How technology can improve communication and health outcomes in patients with advanced cancer: an integrative review. Support Care Cancer 2022 Aug 12;30(8):6525-6543. [doi: 10.1007/s00520-022-07037-y] [Medline: 35411467]
101. Sieck CJ, Rastetter M, Hefner JL, Glover AR, Magaña C, Gray DM, et al. The five A's of access for TechQuity. J Health Care Poor Underserved 2021;32(2S):290-299. [doi: 10.1353/hpu.2021.0064]
102. Rivera V, Aldridge MD, Ornstein K, Moody KA, Chun A. RESEARCHRacial and socioeconomic disparities in access to telehealth. J Am Geriatr Soc 2021 Jan 24;69(1):44-45 [FREE Full text] [doi: 10.1111/jgs.16904] [Medline: 33075143]
103. Benda NC, Veinot TC, Sieck CJ, Ancker JS. Broadband internet access is a social determinant of health!. Am J Public Health 2020 Aug;110(8):1123-1125. [doi: 10.2105/ajph.2020.305784]
104. Zahnd WE, Bell N, Larson AE. Geographic, racial/ethnic, and socioeconomic inequities in broadband access. J Rural Health 2022 Jun 18;38(3):519-526. [doi: 10.1111/jrh.12635] [Medline: 34792815]
105. Majeed A, Maile E, Coronini-Cronberg S. Covid-19 Is Magnifying the Digital Divide. URL: https://blogs.bmj.com/bmj/2020/09/01/covid-19-is-magnifying-the-digital-divide/ [accessed 2022-07-07]
106. National Cancer Institute at the National Institutes of Health. Cancer Disparities. URL: https://tinyurl.com/mryz4846 [accessed 2022-07-07]

107. Vohra J, Marmot MG, Bauld L, Hiatt RA. Socioeconomic position in childhood and cancer in adulthood: a rapid-review. J Epidemiol Community Health 2016 Jun 29;70(6):629-634 [FREE Full text] [doi: 10.1136/jech-2015-206274] [Medline: 26715591]

## Abbreviations

**IHC:** Institute for Healthcare Improvement
**IPV:** in-person visit
**IRB:** institutional review board
**NCI:** National Cancer Institute
**VCV:** videoconferencing visit

---

*Edited by G Eysenbach; submitted 27.05.22; peer-reviewed by K Turner, E Donaghy; comments to author 17.06.22; revised version received 08.07.22; accepted 29.07.22; published 08.09.22.*

*Please cite as:*
*Koppel PD, De Gagne JC, Docherty S, Smith S, Prose NS, Jabaley T*
*Exploring Nurse and Patient Experiences of Developing Rapport During Oncology Ambulatory Care Videoconferencing Visits: Qualitative Descriptive Study*
*J Med Internet Res 2022;24(9):e39920*
*URL: https://www.jmir.org/2022/9/e39920*
*doi:10.2196/39920*
*PMID:36074558*

Original Paper

# Emergency Visits and Hospitalization After Chat Message, Voice Call, or Video Call for Telehealth in Obstetrics and Gynecology Using Telehealth Service User Data in Japan: Cross-sectional Study

Koichi Sakakibara[1], BA; Daisuke Shigemi[2], MPH, MD; Rena Toriumi[2], MD; Ai Ota[2], PHN; Nobuaki Michihata[3], MD; Hideo Yasunaga[4], MD, PhD

[1]Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Kids Public Inc, Tokyo, Japan

[3]Department of Health Services Research, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[4]Department of Clinical Epidemiology and Health Economics, School of Public Health, The University of Tokyo, Tokyo, Japan

**Corresponding Author:**
Koichi Sakakibara, BA
Department of Health, Behavior, and Society
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
Baltimore, MD, 21205
United States
Phone: 1 03 4405 9862
Fax: 1 03 4405 9862
Email: ksakaki1@jhmi.edu

## Abstract

**Background:** In obstetric and gynecologic practices, synchronous telehealth services via chat message, voice calls, and video calls have been increasingly equipped to improve patients' health care accessibility and clinical outcomes. Nevertheless, differences in clinical outcomes between communication tools remain unknown, especially in terms of safety.

**Objective:** This study compared the occurrence of emergency visits and hospitalization after telehealth services through different communication tools, including chat messages, voice calls, and video calls.

**Methods:** We collected data on obstetric and gynecologic concerns of women who consulted specialized doctors and midwives through a telehealth consulting service in Japan (Sanfujin-ka Online) between January 1, 2019, and December 31, 2020. The outcomes were emergency visits or hospitalizations at night after the consultation. Chi-square test and multivariate logistic regression analysis were performed to compare the clinical outcomes between the groups who received telehealth services via chat message, voice calls, and video calls.

**Results:** This study included 3635 participants. The mean age of the participants was 31.4 (SD 5.7) years, and the largest age group (n=2154, 59.3%) was 30-39 years. The numbers (or proportions) of those who received telehealth services via chat message, voice calls, and video calls were 1584 (43.5%), 1947 (53.6%), and 104 (2.9%), respectively. The overall incidence of the outcome was 0.7% (26/3635), including 10 (0.3%) cases of chat message, 16 (0.5%) cases of voice calls, and no video calls. There were no emergency visits that happened due to inappropriate advice. No significant difference in the proportions of the outcomes was observed between the communication tools ($P$=.55). The multivariate logistic regression analysis showed no significant differences in the outcome between those who used chat message and those who used voice calls (odds ratio 1.63, 95% CI 0.73-3.65).

**Conclusions:** The communication tools of telehealth services in obstetrics and gynecology did not show a significant difference in terms of emergency visits or hospitalizations after using the service.

***(J Med Internet Res 2022;24(9):e35643)*** doi:10.2196/35643

**KEYWORDS**

eHealth; gynecology; chat message; mobile health; mHealth; obstetrics; safety; telehealth; telemedicine; video call; voice call

# Introduction

Telehealth refers to various health care services that use digital information and communication technologies. They are categorized into synchronous or asynchronous approaches; synchronous approaches comprise real-time virtual communication and monitoring by health care professionals, whereas asynchronous approaches, also called store-and-forward telehealth, use electronic recording and screening of patient-generated health data [1]. The World Health Organization has highlighted telehealth's potential to improve patients' access to health care services and reduce medical costs [2]. The demand for telehealth services has been expanding, especially in response to the COVID-19 pandemic and physical movement restrictions [3].

In obstetric and gynecologic practices, telehealth has been increasingly equipped to improve patients' accessibility to health care services. Past studies have demonstrated the effectiveness of telehealth services in obstetrics and gynecology, including an improvement in health behaviors [4,5], weight management [6], mental health issues [7], and postpartum depression [8] in pregnant women. Another review has reported the association between telehealth and the improvement of health outcomes in high-risk obstetric patients, such as women with hypertensive disorders, diabetes mellitus, fetal anomalies, and pregnancy in underserved areas [9].

Despite the accumulated knowledge of telehealth in obstetrics and gynecology, scant attention has been paid to safety per the service characteristics. It is especially critical to closely evaluate the safety of synchronous telehealth services, as medical practitioners directly intervene in patients' medical behaviors, while communication via synchronous telehealth tools may limit the information that a medical practitioner can obtain from a patient. The existing literature has reported no significant difference in health outcomes between synchronous telehealth services and traditional in-person services. For example, the Antenatal and Neonatal Guidelines, Education, and Learning System, one of the largest telehealth obstetric care programs providing a phone consultation service for pregnant women in the United States, was as effective as traditional in-person services [10,11]. Text-4-Baby is another telehealth service for pregnant women, and its text messaging service was associated with an improvement in women's health behaviors and attitudes toward child-rearing [12,13]. Additionally, programs with synchronous videoconferences were not associated with increased risks of health complications and medical adherence in comparison with traditional outreach programs [14,15].

The remaining knowledge gap is the comparison between communication tools. Each communication tool (ie, chat message, voice calls, and video calls) may provide different amounts and quality of patients' information to medical practitioners, which can result in divergence in interventions and health outcomes. Thus, we analyzed the secondary data of an online synchronous medical consultation service to compare and evaluate the safety of communication tools (ie, chat message, voice calls, and video calls). Although there are no established safety indicators for online consultations, we defined emergency visits and hospitalization as postconsultation clinical outcomes because emergency hospital visits resulting from inappropriate advice at the time of telehealth services are an important clinical concern.

This study analyzed data from Kids Public Inc, a Japanese health care company providing an online health consultation service in obstetrics and gynecology (Sanfujin-ka Online). The service allows women to consult specialized doctors and midwives about obstetric and gynecologic concerns. Women can use this consultation service at any time before and during pregnancy as well as after childbirth. In Japan, there are no restrictions or regulations regarding the provision or use of telehealth consultations based on the weeks of pregnancy. The users complete a medical questionnaire before each consultation, and medical professionals monitor the urgency based on their responses and the consultations. This service provides only consultations with medical professionals and does not offer medical services such as diagnosis or prescription. Therefore, apart from questions regarding symptoms and medical examinations, the service receives consultations regarding small concerns and questions about health-related daily life issues. In total, 171 medical professionals (ie, obstetrician-gynecologists and midwives) were registered as consultants and responded to consultation requests from 6 PM to 10 PM on weekdays. As Kids Public Inc mainly works with corporations and local governments, most users can avail of the service without payment. This service was approved as a commissioned project by the Ministry of Economy, Trade and Industry in Japan in May 2020 and provided free service to all Japanese citizens until the end of August 2020.

# Methods

## Recruitment and Data Description

We collected data from Kids Public Inc users who responded to a survey between January 1, 2019, and December 31, 2020. Kids Public Inc developed the web-based survey, and some authors were involved in its development as members of Kids Public Inc. This voluntary survey was sent to the users via email automatically within 24 hours of the consultation. The survey was password protected, and service users could carry out data entry via the internet. Other than this one-time self-reporting survey, no protocols for monitoring the patients' postconsultation health behaviors and hospital visits were implemented. There were no incentives for respondents, but they were allowed to skip the questions if they were not comfortable with answering. The data collected through the survey contained the communication tools for consultation (ie, chat message, voice call, or video call), emergency visits or hospitalization at night after the consultation, the consultant category (ie, obstetrician or midwife), the user's situation at the time of consultation (eg, pregnancy, post partum, or other), and whether the hospital visit was unexpected. Data on users' age were also collected because several studies on telehealth services during the pandemic identified that older age was associated with lower use of digital health services [16,17]. Although the demographic distribution of telehealth service use has not been closely explored in Japan, the government report in 2021 addressed that age disparities may have been critical in telehealth use [18]. Consent to use the data anonymously was obtained from all participants at the time of consultation.

### Statistical Analysis

We conducted a secondary analysis of anonymized data to examine differences in the proportion of emergency visits or hospitalizations after consultations according to the communication tools (ie, chat message, voice call, or video call). The exposures, in this study, were the 3 types of communication tools, and the outcomes were the emergency visits or hospitalizations at night after the consultation. Further, consultation records were reviewed to assess whether the emergency visits were unexpected and whether they were caused by inappropriate advice. A chi-square test was used to compare the proportions of the groups. Additionally, multivariate logistic regression analysis was performed with the outcome as the objective variable. Other acquired information was used as a covariate to adjust the users' background. Odds ratios (ORs) and 95% CIs were calculated. The model's goodness of fit was confirmed using the Hosmer-Lemeshow test. All values included in the multivariate analysis were evaluated for correlations, and the absence of multicollinearity was confirmed.

All statistical analyses were performed using Stata software (version 16.0; StataCorp LP). All 95% CIs and *P* values were based on 2-sided hypothesis tests, where *P*<.05 was considered to denote statistical significance.

### Ethics Approval

This study was approved by the Institutional Review Board of the University of Tokyo for joint research between Kids Public Inc and the University of Tokyo (number 2020043NI). Two faculty members of the Department of Clinical Epidemiology and Health Economics at the Graduate School of Medicine, University of Tokyo, conducted and confirmed the analyses to ensure the study's neutrality and transparency. Moreover, CHERRIES checklist was followed, as this is a useful guideline for investigators reporting results of web surveys [19].

## Results

We collected 3635 responses from the web-based survey after the consultations, and the response rate was 40.9%. Table 1 shows the characteristics of the eligible participants. As for the communication tools, voice call was the most common (n=1947, 53.6%), followed by chat message (n=1584, 43.5%). Video call was used by a small proportion of participants (n=104, 2.9%). Overall, the mean age was 31.4 (SD 5.7) years, and the largest age group (n=2154, 59.3%) was 30-39 years. Women under 30 years of age tended to choose chat message. During the consultation, more than half of the participants (n=1883, 51.8%) were in the postpartum period, followed by participants during pregnancy (n=1062, 29.2%), and others (n=690, 19%). Women in postpartum period were the most common users of each tool, but pregnant women were relatively more common in the chat message group. Approximately 60% (n=2174) of the participants consulted obstetrician-gynecologists. In the chat message group, most consultations were with doctors, whereas, in the video call group, most consultations were with midwives. In the voice call group, the proportion of doctors and midwives was almost equal.

Table 2 shows the incidence of primary outcomes divided by communication tools. Of the 3635 responses, 26 (0.7%) cases were reported, including 16 (0.5%) emergency hospital visits and 10 (0.3%) emergency hospitalizations. Of these 26 cases, 10 (0.7%) used chat message, and 16 (0.8%) used voice calls. No outcome incidence was observed among participants using video calls. No significant differences in outcome incidence between communication tools were observed (*P*=.55).

**Table 1.** Participants' characteristics (N=3635).

| Variables | Total (N=3635), n (%) | Chat message (n=1584, 43.5), n (%) | Voice call (n=1947, 53.6), n (%) | Video call (n=104, 2.9), n (%) | *P* value |
|---|---|---|---|---|---|
| **Age (years)** | | | | | <.001 |
| <20 | 130 (3.6) | 73 (4.6) | 54 (2.8) | 3 (2.9) | |
| 20-29 | 1103 (30.3) | 517 (32.6) | 564 (29) | 21 (20.2) | |
| 30-39 | 2154 (59.3) | 905 (57.1) | 1178 (60.5) | 72 (69.2) | |
| ≥40 | 248 (6.8) | 89 (5.6) | 151 (7.8) | 8 (7.7) | |
| **Perinatal situation** | | | | | <.001 |
| Pregnant | 1062 (29.2) | 533 (33.7) | 518 (26.6) | 11 (10.6) | |
| Post partum | 1883 (51.8) | 712 (44.9) | 1088 (55.9) | 83 (79.8) | |
| Other | 690 (19) | 339 (21.4) | 341 (17.5) | 10 (9.6) | |
| **Consultant** | | | | | <.001 |
| Doctor (obstetrician-gynecologist) | 2174 (59.8) | 1129 (71.3) | 1006 (51.7) | 39 (37.5) | |
| Midwife | 1461 (40.2) | 455 (28.7) | 941 (48.3) | 65 (62.5) | |

**Table 2.** Incidence of emergency night visits or hospitalizations within 24 hours after the consultation via communication tools (N=26).

| Variables | Participants, n (%) | *P* value |
|---|---|---|
| **Communication tool** | | .55 |
| Chat message | 10 (0.7) | |
| Voice call | 16 (0.8) | |
| Video call | 0 (0) | |

Table 3 shows the results of the multivariate logistic regression analysis of the primary outcomes. No significant difference in the OR for outcome incidence was observed between chat message and voice calls. Video calls were not included in the regression analysis because the number of outcome occurrences was zero. No significant association was shown with age (OR 0.97, 95% CI 0.90-1.05), consultant occupation (OR 0.63, 95% CI 0.17-2.31), or postpartum period (OR 0.57, 95% CI 0.11-2.96) at the time of consultation, while pregnancy status was significantly associated with the outcome (OR 3.54, 95% CI 1.11-11.3).

In all cases that resulted in emergency visits or hospitalizations, the consulted physician or midwife explained the necessity of emergency visits or seeing a doctor when symptoms worsened, which meant there were no emergency visits or hospitalizations.

**Table 3.** The multivariate logistic regression analysis of emergency night visits or hospitalizations within 24 hours after consultation via communication tools.

| Variables | OR[a] (95% CI) |
|---|---|
| Age[b] (years) | 0.968 (0.869-1.05) |
| **Communication tool** | |
| Chat message | Ref[c] |
| Voice call | 1.63 (0.731-3.65) |
| **Perinatal situation** | |
| Pregnant | 3.54 (1.11-11.3) |
| Post partum | 0.57 (0.11-2.96) |
| Other | Ref (1.08-2.03) |
| **Consultant** | |
| Doctor (obstetrician-gynecologist) | Ref |
| Midwife | 0.63 (0.17-2.31) |

[a] OR: odds ratio.

[b] Continuous variable.

[c] Ref: reference group.

## Discussion

### Principal Results

This study was designed to investigate the safety of online consulting services in obstetrics and gynecology using communication tools (ie, chat message, voice calls, and video calls). Of 3635 collected samples from an online consultation service, 26 (0.7%) cases of emergency visits or hospitalization at night after consultation use were reported. There was no significant difference in the primary outcome incidence between communication tools. Additionally, the results of the multivariate logistic regression analysis indicated no significant differences in the OR of outcome incidence between chat message and voice calls, even when adjusting for the participants' age, consultants' occupation, and participants' situation at the time of consultation. The findings of this study suggest that there are no significant differences in clinical safety among communication tools. However, video calls could not be evaluated in the regression analysis because the number of outcome occurrences was zero. Moreover, the low rate of hospital visits after service use could be attributed to the characteristics of the service; this telehealth service provides only consultations with medical professionals and does not offer medical services such as diagnosis or prescription, so the main users tend to be women who have small concerns and daily-life issues rather than those who have visible symptoms that potentially require hospital visits. Although the service monitors the level of urgency through the completion of a questionnaire before service use and online consultation with a specialist, this finding suggest that the monitoring protocols and method of evaluating consultations after service use may need to be improved. It is important to design service improvements and further research based on this finding.

In this study, users were allowed to choose one of three communication tools, and most participants selected either voice calls (n=1947, 53.6%) or chat message (n=1584, 43.5%). The proportion of video calls was very low (n=104, 2.9%). These results indicate the importance of setting up remote consultation services equipped with a variety of communication tools. In the field of obstetrics and gynecology, video calls were not favored by users, suggesting low compatibility with online consultations. In addition, differences in user background were observed among the three tools: young people under 30 years and pregnant women chose chat message more often. From this perspective, it is preferable to have a variety of communication tools.

The novelty of our study is that we have shown the clinical safety of online obstetric and gynecologic consulting services. Despite the emerging knowledge about synchronous telehealth services, few studies have addressed the clinical safety of telehealth services using multiple communication tools like chat message and voice calls. Past studies have analyzed telehealth programs without categorizing them by the nature of the program (ie, synchronous and asynchronous) or communication tools [10-15]. However, it is worth noting that studies with a single communication tool reported no clinical differences between traditional and telehealth approaches, which is logically consistent with the findings in our study. Nevertheless, as video calls were associated with only approximately 100 cases and no outcomes, adequate evaluation could not be conducted using regression analysis in this study. We hope to reevaluate the safety of video calls when more data are available.

## Limitations

The generalizability of this study is subject to certain limitations, owing to the characteristics of the study population. The age distribution of the participants was slightly different from that of the nationwide population. According to a recent government report, the age distribution of pregnant women in Japan is as follows: 0.83% for those under 20 years of age, 33.84% for those between 20 and 29 years, 59.44% for those aged 30-39 years, and 5.90% for those aged 40 years or older [20]. Compared to the national distribution, the study participants had a larger proportion of people aged under 20 years and between 20 and 29 years. This may be because using online services are more common among the younger generation. Additionally, the service was funded by the government and became available for those who were not members of client organizations between May and August 2020. This could have influenced the user demographics during the study period because the awareness about the service might have increased among the younger generation, who are less likely to work in a corporation or be involved in pregnancy and childbirth. Thus, the findings are limited to this telehealth service and may not be generalized to other telehealth services.

Other limitations of this study need to be acknowledged, particularly in data collection. First, the survey response rate was 40.9%, which may be insufficient to evaluate the behavioral patterns of all service users. Second, due to the small number of video call consultation cases, the analysis could have insufficient statistical power. Third, users’ selection of communication tools may be subject to self-selection bias. The very low number of users who chose video calls may be explained by users’ circumstances during the call, such as not having sufficient equipment or calling while on the go, when they could not use the camera. Fourth, although the survey queried participants’ emergency visits and hospitalization at night after service use, the outcomes may contain hospital visits that are not technically emergencies, which may have led to overestimation. Fifth, the survey questions were not validated to ensure a precise record of patients’ hospital visits after service use. Sixth, we could not analyze multiple hospital visits after service use because neither the survey items nor other monitoring protocols tracked the study participants’ frequency of and reasons for hospital visits after the consultation. Other variables, such as medical history, pregnancy and delivery situations, gestational weeks of pregnancy, socioeconomic status, and family environment could also not be adjusted for this study due to the lack of data. In addition, the final pregnancy outcomes could not be evaluated in this study. Therefore, further research needs to examine the clinical outcomes of telehealth services more closely and with larger data sets, including these contextual factors and the course of pregnancy. Specifically, the use of validated questionnaires and monitoring protocols to document detailed information regarding patients’ hospital visits after service use is encouraged.

## Conclusions

Our analysis suggests that different communication tools for telehealth services in obstetrics and gynecology may not be associated with clinical safety among service users. However, there were several limitations, and the results require interpretation in light of the characteristics of telehealth service provided to the participants. Future research should analyze the data with more emergency cases and relevant variables to examine the consequences of synchronous telehealth consultation services.

---

## Acknowledgments

The authors would like to thank all obstetricians and midwives affiliated with Kids Public Inc and all study participants for their contributions to the study.

---

## Conflicts of Interest

DS, RT, and AO were employed by Kids Public Inc, Japan. Other authors report no conflicts of interest.

---

## References

1. Brown HL, DeNicola N. Telehealth in maternity care. Obstet Gynecol Clin North Am 2020 Sep;47(3):497-502. [doi: 10.1016/j.ogc.2020.05.003] [Medline: 32762934]
2. WHO Global Observatory for eHealth. Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth. Geneva, Switzerland: World Health Organization; 2010:1-93.
3. Demeke HB, Merali S, Marks S, Pao LZ, Romero L, Sandhu P, et al. Trends in use of telehealth among health centers during the COVID-19 pandemic - United States, June 26-November 6, 2020. MMWR Morb Mortal Wkly Rep 2021 Feb 19;70(7):240-244 [FREE Full text] [doi: 10.15585/mmwr.mm7007a3] [Medline: 33600385]
4. DeNicola N, Grossman D, Marko K, Sonalkar S, Butler Tobah YS, Ganju N, et al. Telehealth interventions to improve obstetric and gynecologic health outcomes: a systematic review. Obstet Gynecol 2020 Feb;135(2):371-382 [FREE Full text] [doi: 10.1097/AOG.0000000000003646] [Medline: 31977782]
5. Whittington JR, Ramseyer AM, Taylor CB. Telemedicine in low-risk obstetrics. Obstet Gynecol Clin North Am 2020 Jun;47(2):241-247. [doi: 10.1016/j.ogc.2020.02.006] [Medline: 32451015]
6. Sherifali D, Nerenberg KA, Wilson S, Semeniuk K, Ali MU, Redman LM, et al. The effectiveness of eHealth technologies on weight management in pregnant and postpartum women: systematic review and meta-analysis. J Med Internet Res 2017 Oct 13;19(10):e337 [FREE Full text] [doi: 10.2196/jmir.8006] [Medline: 29030327]
7. Worley LLM, Wise-Ehlers A. Telepsychiatry in obstetrics. Obstet Gynecol Clin North Am 2020 Jun;47(2):333-340. [doi: 10.1016/j.ogc.2020.02.009] [Medline: 32451021]
8. Chow R, Huang E, Li A, Li S, Fu SY, Son JS, et al. Appraisal of systematic reviews on interventions for postpartum depression: systematic review. BMC Pregnancy Childbirth 2021 Jan 06;21(1):18 [FREE Full text] [doi: 10.1186/s12884-020-03496-5] [Medline: 33407226]
9. Whittington JR, Magann EF. Telemedicine in high-risk obstetrics. Obstet Gynecol Clin North Am 2020 Jun;47(2):249-257. [doi: 10.1016/j.ogc.2020.02.007] [Medline: 32451016]
10. Ivey TL, Hughes D, Dajani NK, Magann EF. Antenatal management of at-risk pregnancies from a distance. Aust N Z J Obstet Gynaecol 2015 Feb;55(1):87-89. [doi: 10.1111/ajo.12254] [Medline: 25355283]
11. Britt DW, Norton JD, Hubanks AS, Navidad SA, Perkins RJ, Lowery CL. A two-period assessment of changes in specialist contact in a high-risk pregnancy telemedical program. Telemed J E Health 2006 Feb;12(1):35-41. [doi: 10.1089/tmj.2006.12.35] [Medline: 16478411]
12. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012;12:1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
13. Huberty JL, Buman MP, Leiferman JA, Bushar J, Hekler EB, Adams MA. Dose and timing of text messages for increasing physical activity among pregnant women: a randomized controlled trial. Transl Behav Med 2017 Jun;7(2):212-223 [FREE Full text] [doi: 10.1007/s13142-016-0445-1] [Medline: 27800565]
14. Pflugeisen BM, McCarren C, Poore S, Carlile M, Schroeder R. Virtual Visits: managing prenatal care with modern technology. MCN Am J Matern Child Nurs 2016;41(1):24-30. [doi: 10.1097/NMC.0000000000000199] [Medline: 26474477]
15. Rhoads SJ, Serrano CI, Lynch CE, Ounpraseuth ST, Gauss CH, Payakachat N, et al. Exploring implementation of m-Health monitoring in postpartum women with hypertension. Telemed J E Health 2017 Oct;23(10):833-841 [FREE Full text] [doi: 10.1089/tmj.2016.0272] [Medline: 28475431]
16. 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 Mar 01;147(3):287-295 [FREE Full text] [doi: 10.1001/jamaoto.2020.5161] [Medline: 33443539]
17. 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 Oct;23(5):368-377. [doi: 10.1089/pop.2020.0186] [Medline: 32816644]
18. White paper 2021. Ministry of Internal Affairs and Communications, Japan. URL: https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/whitepaper/index.html [accessed 2022-09-08]
19. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Dec 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Summary of vital statistics in 2020 (in Japanese). Ministry of Health, Labour, and Welfare. URL: https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei20/dl/16_all.pdf [accessed 2022-09-08]

## Abbreviations

**OR:** odds ratio

*Edited by G Eysenbach; submitted 29.12.21; peer-reviewed by N Maglaveras, J Su, KL Mauco; comments to author 22.04.22; revised version received 14.06.22; accepted 31.08.22; published 23.09.22.*

*Please cite as:*
*Sakakibara K, Shigemi D, Toriumi R, Ota A, Michihata N, Yasunaga H*
*Emergency Visits and Hospitalization After Chat Message, Voice Call, or Video Call for Telehealth in Obstetrics and Gynecology Using Telehealth Service User Data in Japan: Cross-sectional Study*
*J Med Internet Res 2022;24(9):e35643*
*URL: https://www.jmir.org/2022/9/e35643*
*doi:10.2196/35643*
*PMID:36149744*

Original Paper

# Development of Principles for Health-Related Information on Social Media: Delphi Study

Emily Denniss[1], BNutrSc (Hons); Rebecca Lindberg[1], PhD; Sarah A McNaughton[1], PhD

[1]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Melbourne, Australia

**Corresponding Author:**
Emily Denniss, BNutrSc (Hons)
Institute for Physical Activity and Nutrition
School of Exercise and Nutrition Sciences
Deakin University
221 Burwood Hwy
Melbourne, 3125
Australia
Phone: 61 3 924 68761
Email: e.denniss@deakin.edu.au

## Abstract

**Background:** Health-related misinformation can be propagated via social media and is a threat to public health. Several quality assessment tools and principles to evaluate health-related information in the public domain exist; however, these were not designed specifically for social media.

**Objective:** This study aims to develop Principles for Health-related Information on Social Media (PRHISM), which can be used to evaluate the quality of health-related social media content.

**Methods:** A modified Delphi approach was used to obtain expert consensus on the principles and functions of PRHISM. Health and social media experts were recruited via Twitter, email, and snowballing. A total of 3 surveys were administered between February 2021 and May 2021. The first survey was informed by a literature review and included open-ended questions and items from existing quality assessment tools. Subsequent surveys were informed by the results of the proceeding survey. Consensus was deemed if ≥80% agreement was reached, and items with consensus were considered relevant to include in PRHISM. After the third survey, principles were finalized, and an instruction manual and scoring tool for PRHISM were developed and circulated to expert participants for final feedback.

**Results:** A total of 34 experts consented to participate, of whom 18 (53%) responded to all 3 Delphi surveys. In total, 13 principles were considered relevant and were included in PRHISM. When the instructions and PRHISM scoring tool were circulated, no objections to the wording of the final principles were received.

**Conclusions:** A total of 13 quality principles were included in the PRHISM tool, along with a scoring system and implementation tool. The principles promote accessibility, transparency, provision of authoritative and evidence-based information and support for consumers' relationships with health care providers. PRHISM can be used to evaluate the quality of health-related information provided on social media. These principles may also be useful to content creators for developing high-quality health-related social media content and assist consumers in discerning high- and low-quality information.

***(J Med Internet Res 2022;24(9):e37337)*** doi:10.2196/37337

**KEYWORDS**

social media; health information; information quality; quality assessment; research tool; credibility indicator; credibility; credible; eHealth; Delphi; social media; assessment tool; measurement tool; web-based information; mobile phone

## Introduction

### Background

Health information–seeking behavior can influence an individual's decision-making and overall health [1]. The internet is a popular source of health-related information for the general public [2-5], and its popularity has been growing [6]. The internet has facilitated web-based environments where information can be published and accessed with tremendous speed and ease [7]. Technological innovations have led to the widespread use of devices such as smartphones, tablets, and computers, especially in high- and middle-income countries, and consumer access to the internet and information is now ubiquitous [3,8,9]. Consequently, web-based environments have become highly accessible and efficient sources of health-related information.

Social media is a prominent component of web-based environments. Social media refers to internet-based applications that facilitate user-generated content, allow individuals to create user profiles and identities, and develop web-based networks by connecting user profiles and groups [10]. Each social media platform has unique characteristics; however, all platforms share these common features [10]. Users can amass large followings, and information in the form of text, images, and videos can be instantaneously published and viewed by many people [10,11]. Approximately half of the global population comprises active social media users, and rates of social media use are steadily and continuously increasing [12]. As social media has grown in popularity, so too has its use by consumers for health-related information [13,14]. Consumers not only actively seek health-related information on social media but are also passively exposed to it in their social media feeds [15,16].

The spread of health-related misinformation on social media has been identified as a serious threat to public health [17-19]. All social media users have the freedom to publish and share information on almost any topic, regardless of their credentials, and consequently, information on social media is of variable quality and veracity. In addition, previous research has identified low levels of media literacy among consumers [20], which is the ability to access, understand, and critically evaluate information presented in the media [21]. These factors have combined and contributed to the propagation of health-related misinformation on social media, which has the power to undermine credible public health messaging [17,18]. For example, it is believed that the publication and spread of misinformation on social media have amplified and accelerated the transmission of the SARS-CoV-2 virus [18,22], contributing to decreased vaccination rates and the re-emergence of previously eradicated diseases such as the measles [23,24]. Communication of high-quality health information via social media has potential benefits, such as increased accessibility of information, improved communication between health professionals and patients, and social and emotional support for patients [25]. However, social media health information is generally not of high quality, particularly in relation to cancer, diabetes, and dental care [26].

The quality of health-related information refers to the reliability of information compared against a set of defined quality criteria [27]. There is a large body of literature on health information quality, and many tools and principles have been developed to evaluate the quality of health-related information in specific contexts [27]. The DISCERN instrument [28,29], *Journal of the American Medical Association* (JAMA) benchmarks [30], and the Health on the Net (HONcode) principles [31] are most commonly used to evaluate internet-based health-related information [27]. The DISCERN instrument was established to judge the quality of written health-related information [28], and the JAMA benchmarks and HONcode principles were designed to evaluate and guide the development of information provided on websites [30,31]. These tools were developed by experts, have been extensively used throughout the literature, and have been shown to be reliable and valid measures of information quality across a wide range of health-related topics [27,28,30,31].

Existing quality assessment tools such as the DISCERN instrument, JAMA benchmarks, and HONcode principles and other established tools share common criteria to evaluate health information. Criteria to assess the disclosure of advertising policies, sponsorships, and financial conflicts of interest are included in the most frequently used tools for evaluating internet-based health information [27]. Similarly, existing tools for websites assess whether the date of the information's publication and last update have been disclosed and whether references to the original sources of information have been included [27]. The commonalities between these tools demonstrate the agreement in the literature regarding the fundamental elements of high-quality health information. Therefore, there is now a need to evaluate the quality of health-related social media content. Thus far, in the literature, studies that have evaluated the quality of social media content have used quality assessment tools designed for different contexts [26]. For example, the DISCERN instrument, designed for written information, and HONcode principles, designed for websites, have both been used to evaluate the quality of YouTube and Facebook content [26]. However, the use of these tools to evaluate social media content may not be suitable, given the unique characteristics of social media and that quality principles from existing tools require adaptation to suit social media.

The widely used DISCERN instrument, JAMA benchmarks, and HONcode principles were developed between 1996 and 1998 and have since undergone minor revisions or have not been updated at all [28-32]. To illustrate some of their limitations in the social media setting, consider that content on social media is usually kept brief to increase user engagement and that some platforms place limits on the number of characters, images, or length of videos in posts [33]. The existing quality assessment tools assume that there are no limits to the length or amount of information provided. As such, it is unlikely that health-related social media content can comprehensively address the quality criteria outlined in these existing tools. In addition, disclosure of funding and conflicts of interest are emphasized in prominent quality assessment tools [28-31]. However, the operationalization of the principles of financial disclosure in previous tools does not consider the covert advertising and influencer marketing that exist on social media and, therefore, may not be sensitive enough to be applied to social media.

The DISCERN instrument, JAMA benchmarks, HONcode principles, and similar tools were developed with static information environments in mind, such as patient information pamphlets, websites, and books. In such environments, the public searches for information to consume it, and experts are better able to act as gatekeepers for credible information [32]. Conversely, in dynamic social media environments, users often consume information passively, and there is an emphasis on user-generated content, which blurs the boundaries between information producers and consumers [32]. As a result, it is more challenging to discern authoritative sources, and previous methods of judging a source's credibility may lack relevance in the context of social media.

### Objective

There is an increasing number and diversity of social media platforms, and the use of social media for health information–seeking is also increasing. The quality of health information can affect public health communication in both positive and negative directions. As such, the need to measure the quality of health-related social media content has become increasingly important in research settings. Thus far, no quality assessment tool has been developed to suit the particular context of social media. Furthermore, there is a scarcity of literature that has outlined standards for high-quality health-related social media content that can be used to inform the development and selection of credible content. The research gap regarding information quality on social media has also been identified in 2 systematic reviews, which have both described the need for a suitable tool to evaluate the quality of social media content [25,26]. Therefore, the aim of this study was to develop Principles for Health-related Information on Social Media (PRHISM), which defines high-quality health-related information and can be used to evaluate the quality of health-related social media content posted by any public account across all social media channels.

## *Methods*

### Study Design

This study used a modified Delphi technique, which is a group facilitation method that aims to attain consensus among a panel of experts through iterative surveys and controlled feedback [34,35]. This methodology was selected as it is widely used in health research and is appropriate for facilitating decision-making when there is incomplete knowledge, multiple disciplines involved, or a diversity of opinions on the topic of investigation [34,35]. Furthermore, the Delphi method has been shown to be effective for developing new concepts and is a suitable method for establishing definitions for use in research and practice tools [36-38].

The number of surveys, herein referred to as rounds, was determined a priori [35]. A total of 3 rounds are considered optimal in Delphi studies; therefore, 3 rounds were set to gain consensus and develop PRHISM [35]. Qualtrics software (Qualtrics) was used to host all of the surveys in this study, and surveys were administered between February 2021 and May 2021.

### Ethics Approval

This study was approved by the Deakin University Human Research Ethics Committee (HEAG-H 242_2020).

### Participants and Recruitment

Purposive sampling was used to recruit experts in human health or social media with experience using ≥1 social media platform in a professional capacity for health promotion, research dissemination, or representing a health-related organization. This study used multiple recruitment strategies. Advertising via Twitter was conducted through the accounts of each of the authors and their research institutes based in Melbourne, Australia. An invitation to participate was emailed to the communications departments of 19 Australian and 27 global health organizations that were identified as having a social media presence through Twitter. Snowballing was also used with the recruited participants.

All recruitment materials directed interested parties to a web-based survey to determine their eligibility, register their details, and provide informed consent. To be considered eligible, individuals must have had a minimum of 3 years of experience working anywhere in the world as (1) a health professional registered with a professional body, (2) a researcher or an academic in a health-related field, or (3) a communications or social media specialist in a health-related organization. Such experts were selected as they were familiar with health-related social media content and the characteristics of social media platforms because of their professional experience using social media. Furthermore, it is common for academic institutions, professional bodies, and health-related organizations to have policies or position statements regarding ethical and credible communication on social media. The involvement of eligible individuals with such organizations and the professional use of social media indicated that they were likely to be familiar with issues relating to credible and ethical health communication. Recruitment materials and surveys were provided in only English, and participation was voluntary, with no compensation offered.

### Round 1

Round 1 was informed by a review of the literature. Participants were presented with items from 4 tools. The DISCERN instrument, JAMA benchmarks, and HONcode principles were included as they are the most consistently used instruments in research assessing the quality of health-related information on the internet [27]. The Quality Evaluation Scoring Tool (QUEST; developed to assess internet-based health-related information) was also included as it is highly cited, and its more recent development means that it may cover relevant elements that the older tools do not [39]. All 4 tools included items that can be applied to evaluate a source of health information. The source, assessed for accuracy against the items, is ultimately scored to establish its overall quality.

A total of 33 items from the 4 tools were included in round 1. The participants were asked to rate the importance of each item in the context of health-related information provision specifically on social media. Importance was rated on a 5-point Likert scale ranging from *not at all important* to *very important*. Likert scales are highly recommended for rating statements in Delphi studies, and the 5-point scale was chosen as the optimal number of response categories is between 4 and 7 in Delphi research [34,35,40].

Round 1 also included 6 open-ended questions that probed respondents for deeper insights and provided them with the opportunity to provide written comments about the items. At the end of round 1, participants were also asked to suggest principles not already covered in the presented items to include in PRHISM.

### Round 2

Round 2 comprised a second opportunity for rating the original 33 items, with participants having access to the feedback and results from round 1. In addition, participants rated (on the 5-point Likert scale) 9 new principles and 11 comments, both generated from the open-ended questions in round 1. A summary of all verbatim responses to the open-ended questions from round 1 was also provided. Participants were asked to consider the written responses from other group members when responding to the survey. There were no open-ended questions.

Feedback about group responses to Likert scale questions was provided as the median response and IQR of each question. This feedback method for Likert scale questions is consistently recommended in the literature [34,35,40]. Feedback was presented alongside its corresponding question, and as suggested by Trevelyan et al [35], a visual aid, by way of a bar graph, was also provided to assist participants' understanding of how the group responded to the preceding round. In line with best practice Delphi methods [27], participants were asked to consider how the rest of the group responded before rating the importance of the quality assessment items again.

In addition to the existing 33 items, 9 new principles suggested from round 1 were added for round 2. These principles were proposed by participants in response to an open-ended question in round 1, which asked for suggestions of additional principles to be included in PRHISM. Suggestions of new principles were content analyzed following the procedure outlined by Keeney et al [40]. Responses that were the same or very similar were initially merged by the first author (ED) before all the authors met to discuss. Disagreements regarding the merging of principles were discussed among the authors until an agreement was reached. Participants were asked to rate the importance of the new principles in round 2 on the 5-point Likert scale.

In addition to the quality principles (33 original items and 9 new principles), 11 comments were included, and participants were asked to rate their level of agreement on a 5-point Likert scale ranging from *strongly disagree* to *strongly agree*. These comments were generated from the responses to the remaining open-ended questions from round 1, which were content analyzed, and merged where similar, following the same method that was used for the newly suggested principles. Comments about items from round 1, which could be used to inform adaptations to the principles that may ultimately be included in PRHISM, were provided to participants in round 2.

### Round 3

A total of 22 principles were included in round 3, comprising 13 adapted principles, 6 principles from pre-existing quality assessment tools, and 3 new principles suggested by participants in round 1. Principles (ie, the items) from round 2 achieving consensus on being important to include in PRHISM were adapted where necessary and presented to participants in round 3 following a method similar to Mete et al [38]. Principles were adapted to make them relevant to social media and incorporate participant comments that were agreed upon in round 2. Similar principles where consensus was reached in round 2 were merged, as the 4 original tools have comparable aims and applications and, hence, some similar items. The adaptation and merging of principles were initially performed by the lead author (ED). Once updated, principles were circulated to the authors, and disagreements about the changes were discussed until an agreement was reached. Several principles were deemed appropriate and did not need to be amended. These were omitted from round 3 as consensus had already been reached in round 2.

A summary of how the principles had been adapted and merged was provided at the start of round 3. Participants were asked to rate their level of support for the inclusion of the adapted principles in the final PRHISM tool on a 5-point Likert scale ranging from *strongly oppose* to *strongly support*. Principles that did not achieve consensus in round 2 were also included in round 3. Participants were presented with feedback in the same format as in round 2 and asked to consider the groups' responses when rating the importance of these items again.

### Consensus

Consensus was calculated after rounds 2 and 3 for each survey item. It was deemed that consensus had been met if ≥80% of participants selected 4 or 5 (important) or 1 or 2 (not important) on the Likert scale. Although determining consensus varies in Delphi methods [34,35,40], the 80% cutoff point was chosen as it has been suggested as an appropriate figure in some of the Delphi literature [40] and has been used in a similar Delphi study [38]. Stability of responses between survey iterations is often assessed in Delphi studies to aid in determining whether a consensus has been achieved [40,41]. There is a limited agreement in the literature on how to measure the stability of participant responses, and increasing weighted κ values is the most consistently advised method [40,41]. Owing to the adaptation of the principles included in the iterative surveys, it was not possible to calculate the κ values, and no measure of response stability was included. This approach is consistent with a recent Delphi study with aims and methods similar to this study [38].

### Development of PRHISM and Scoring Tool

Principles that reached a consensus were included in the final PRHISM tool. If consensus was not achieved by round 3, it was concluded that the principle was not relevant to be included in PRHISM. For those principles where consensus emerged only in round 3 (and not in previous rounds), the process of adapting and merging the principles was repeated. Once agreement about the adaptation of principles had been reached among the authors, the updated principles were also included in PRHISM.

A scoring system with instructions was also developed, which outlines the criteria to define what can be considered low-, moderate-, or high-quality information. This was based on the handbook and scoring system of the DISCERN instrument, which uses a 5-point Likert scale for rating the degree to which each quality principle has been met [29]. Participant comments where consensus was reached in round 2 were also used to inform the instructions for how each principle should be scored when evaluating health-related information on social media.

After the draft of the PRHISM instructions and scoring tool had been finalized, participants who had completed at least round 1 were invited to provide final feedback. Although not an official survey round, this step enabled communication of the agreed principles and captured any final expert perspectives on the applications and utility of the tool.

## Results

### Overview

A total of 34 eligible experts consented to participate in this study, of whom 26 (76%) completed round 1, a total of 23 (68%) completed round 2, and 18 (53%) participants completed round 3 (69% retention rate from round 1). Participants were from Australia, Italy, China, New Zealand, the United Kingdom, the United States, and Vietnam and had expertise in a range of health-related disciplines (Table 1). Participants who completed all 3 Delphi rounds had an average of 10 (SD 8) years of experience in their health-related discipline, with a minimum of 3 and a maximum of 21 years of experience. The participants used a range of social media platforms (Table 1). Holding a personal social media account for professional purposes, managing the social media account of a health-related professional organization, and creating blog content were reported as participants’ professional health-related uses of social media.

**Table 1.** Participant characteristics across 3 rounds of the Delphi process (N=26).

| Characteristics | Round 1 (n=26), n (%) | Round 2 (n=23), n (%) | Round 3 (n=18), n (%) |
|---|---|---|---|
| **Eligibility criteria met**[a] | | | |
| Researcher or academic | 15 (58) | 14 (61) | 12 (67) |
| Health professional | 10 (38) | 8 (35) | 5 (28) |
| Communications or social media staff | 4 (15) | 4 (17) | 3 (17) |
| **Health-related field**[a] | | | |
| Allied health | 9 (35) | 7 (30) | 5 (28) |
| Public health | 10 (38) | 9 (39) | 8 (44) |
| Medicine and clinical care | 4 (15) | 3 (13) | 2 (11) |
| Preventive health | 14 (54) | 14 (61) | 11 (61) |
| **Highest level of education** | | | |
| Doctorate | 10 (38) | 9 (39) | 7 (39) |
| Master’s degree | 10 (38) | 9 (39) | 8 (44) |
| Graduate diploma | 1 (4) | —[b] | — |
| Bachelor’s degree with Honors | 2 (8) | 2 (9) | 1 (6) |
| Bachelor’s degree | 3 (12) | 3 (13) | 2 (11) |
| **Location** | | | |
| Australia | 20 (77) | 17 (74) | 13 (72) |
| Other[c] | 6 (23) | 6 (26) | 5 (28) |
| **Social media use**[a] | | | |
| Personal account for professional use | 24 (92) | 21 (91) | 17 (94) |
| Management of professional organization’s account | 12 (46) | 10 (43) | 7 (39) |
| Blog writing | 5 (19) | 4 (17) | 2 (11) |
| **Social media platforms used**[a] | | | |
| Facebook | 19 (73) | 17 (74) | 13 (72) |
| LinkedIn | 21 (81) | 18 (78) | 15 (83) |
| Snapchat | 3 (12) | 3 (13) | 2 (11) |
| Instagram | 18 (69) | 16 (70) | 11 (61) |
| YouTube | 9 (35) | 8 (35) | 6 (33) |
| Twitter | 24 (92) | 21 (91) | 17 (94) |
| Pinterest | 1 (4) | 1 (4) | 1 (6) |
| Reddit | 2 (8) | 1 (4) | 1 (6) |
| TikTok | 2 (8) | 1 (4) | 1 (6) |
| Clubhouse | 1 (4) | — | — |

[a] Participants could fall under >1 category or select >1 option.

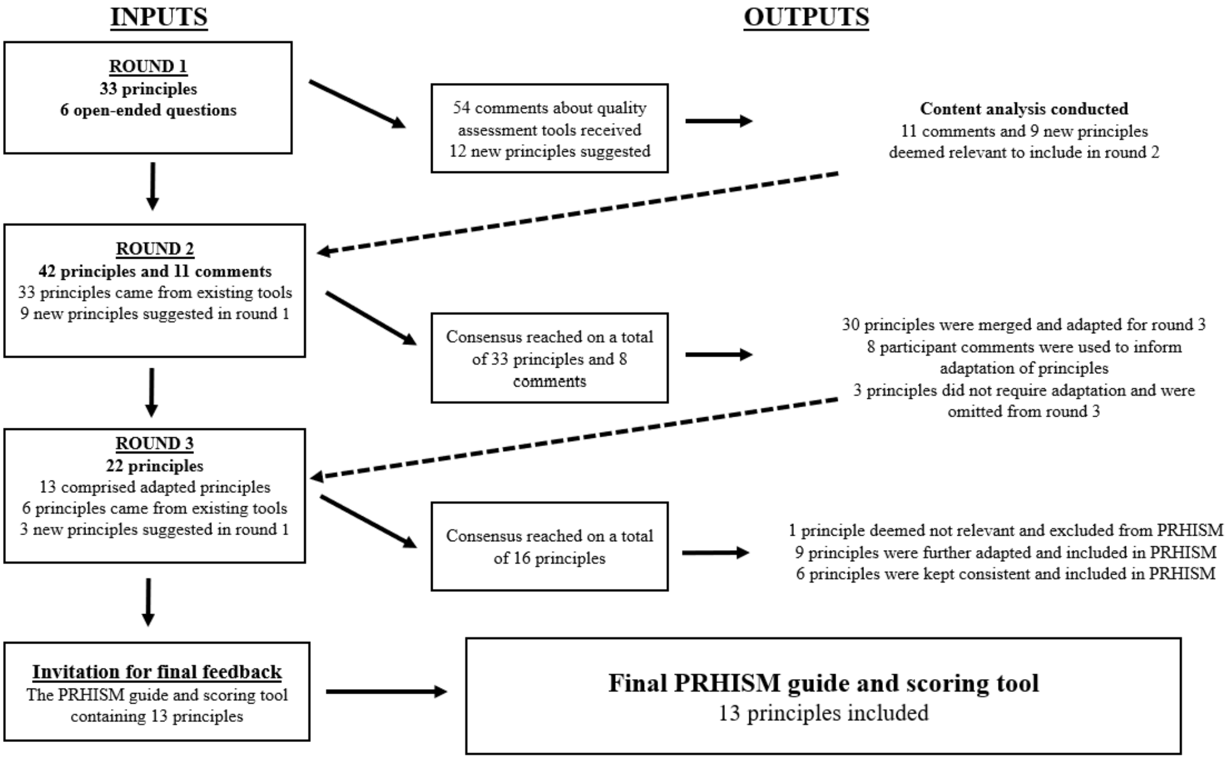
[b] Not available.

[c] One participant from Italy, China, New Zealand, United States, United Kingdom, and Vietnam each.

## Round 1

The Delphi process is summarized in Figure 1 [38]. Participants rated a total of 33 items from 4 pre-existing quality assessment tools and were presented with 6 optional open-ended questions. A total of 54 comments were received in response to the open-ended questions in round 1. Participants also suggested 12 new principles that they considered important to the provision of quality health-related information on social media and that had not already been outlined in the quality assessment tools presented to them. After the content analysis was performed on the open-ended responses from round 1, a total of 11 comments about existing principles and 9 new principles were carried forward to round 2.

**Figure 1.** Flowchart of Delphi process adapted from Mete et al [38]. PRHISM: Principles for Health-related Information on Social Media.



## Round 2

A total of 6 new principles, 27 principles from existing tools, and 8 participant comments reached consensus in round 2. All principles reaching consensus achieved consensus with ≥80% of participants selecting 4 or 5 on the Likert scale (ie, important), and none of the principles achieved consensus for participants selecting 1 or 2 (ie, unimportant). Similarly, all comments with consensus had ≥80% of participants selecting 4 or 5 on the Likert scale (ie, agree). Of the 33 principles that reached consensus, 30 (91%) were updated to incorporate consensus comments and merged where similar to form 13 adapted principles, which were included in round 3. The 3 remaining principles with consensus did not require updating and were thus omitted from round 3. The consensus items from round 2 and their adaptations are summarized in Multimedia Appendix 1.

## Round 3 and Development of PRHISM

Consensus was reached on a total of 16 principles. Of these 16 principles, 10 (62%) were adapted principles, 4 (25%) were principles from existing quality assessment tools, and 2 (13%) were new principles suggested in round 1. Similar to round 2, all principles with consensus in round 3 achieved ≥80% agreement through participants selecting 4 or 5 on the Likert scale (ie, important). One of the items from an existing tool reached consensus in round 3 (83% agreement) but was not included in PRHISM ("Does the information achieve its aims?" DISCERN instrument, question 2) [29]. It was excluded as it depended upon an adapted principle that did not achieve consensus (78% agreement) in round 3 ("The aims of health-related information provided on social media should be clearly outlined." Adapted from the DISCERN instrument, question 1). The research team excluded both items from PRHISM as the importance of stating and achieving aims was only mentioned by 1 of the 4 tools in the surveys. Therefore, it was less likely to be critical to the provision of quality health-related information.

Of the remaining 15 principles that achieved consensus in round 3, a total of 6 (40%) were kept consistent and included in PRHISM, and 9 (60%) were further merged and adapted before being included in PRHISM, including 1 principle, which was merged with a consensus principle from round 2 that was not tested in round 3. The additional 2 principles that achieved consensus in round 2, and thus not tested in round 3, were also included, providing a total of 13 principles in PRHISM (Textbox 1). These 13 principles outline the gold standard for high-quality health-related social media content. The consensus items from round 3 and their adaptations are summarized in Multimedia Appendix 1.

**Textbox 1.** Final principles of the Principles for Health-related Information on Social Media tool.

**Principle and description**

Principle 1: authorship

When providing health-related information on social media, the authors and contributors, their credentials, and their affiliations should be clearly stated on the social media profile. If this information cannot fit on a profile, crediting an authoritative institution is sufficient, if relevant. If not, all the contributors, their credentials, and their affiliations should be included.

Principle 2: authoritative

Health-related information provided on social media should be provided by qualified professionals, including health and medical scientists, and information should be within the scope of practice of the author’s qualifications. If information is provided by an unqualified person, this should be clearly indicated.

Principle 3: action oriented

Health-related information provided on social media should be action oriented and include clear, succinct messages to support decision-making and provide context for the consumer.

Principle 4: financial disclosure

Sponsorship, advertising, funding arrangements, and financial support or any potential conflicts of interest should be fully disclosed in a prominent and clear manner. Financial support and conflicts of interest can be disclosed on the social media profile. However, if a post has been sponsored, paid for, and contains advertising or a product that has been gifted, this needs to be clearly and prominently disclosed in the social media post.

Principle 5: attribution

Health-related information on social media should include clear references and hyperlinks to the original source of information used to compile the post. It should be clear when the original source of information was published. If all references cannot fit into the social media post, a link to the references and further information should be provided.

Principle 6: balance and justifiability

Health-related information provided on social media, which includes claims relating to the benefits or performance of a particular treatment, product, service, or behavior, should be balanced, unbiased, and supported by appropriate and quality evidence. The use of causative language and “shock tactics” should be avoided, and information about limitations or contrasting findings should be included.

Principle 7: risks and benefits

Health-related information provided on social media about a particular treatment, product, service, or behavior should clearly outline associated risks and benefits.

Principle 8: privacy

Health-related information on social media should respect principles of privacy and confidentiality. For example, if information, images, or videos of or about others are shared, they should be shared with permission.

Principle 9: complementary information

Health-related information provided on social media should provide support for individuals’ relationships with their physicians and other professional health care providers and should not be designed to replace such relationships. Support for discussion of options with the individuals’ health care provider should be included in posts containing health-related information.

Principle 10: referrals and support

Health-related information provided on social media should include referrals to additional sources of support and information. Where possible, links to such resources should be included.

Principle 11: readability and comprehensibility

Health-related information on social media should avoid the use of technical language and medical jargon. Plain language should be used, and information should be easily understandable by the general public and written at a grade 5 reading level.

Principle 12: accessibility

Medical and health information provided on social media should be accessible to individuals with vision and hearing impairments. For example, where relevant, social media posts that include images should provide alternative text in the caption, and videos should include closed captions.

Principle 13: images

Images included in health-related social media posts should be visually appealing and reflect rather than contradict the information provided in the post.

## Final Comment Stage

A total of 13 principles and instructions on the operationalization and scoring of each principle was sent to 26 participants. One of the participants responded with minor feedback and questions. No objections to the wording of the principles or instructions were received, and the principles were not adapted any further.

The final PRHISM Guide and Scoring Tool is provided in Multimedia Appendix 2 [39,42-48].

## Discussion

### Principal Findings

This study used a Delphi approach to develop PRHISM. A total of 13 principles were established to define high-quality information and evaluate the quality of health-related social media content. PRHISM can be used to evaluate health-related social media posts intended for nonspecialist audiences, which are posted by public accounts on any social media platform or blog, and can be used as a research tool or to inform the development and selection of high-quality content. Many of the principles included in PRHISM are similar to principles and concepts from other tools [28-31,39]. However, their content and operationalization have been adapted to suit social media. The items in PRHISM can be broadly categorized into 4 themes: accessibility, transparency, authoritative and evidence-based information, and complementary relationships between patients and health professionals. These principles were agreed upon by experts recruited into the Delphi process and are well supported by the literature.

### PRHISM Themes

Accessibility is covered in principles 3, 11, and 12 of PRHISM. Principle 3 stipulates that health-related information should be action oriented, clear, succinct, and facilitative of decision-making. This is in line with evidence that recommends providing practical and simple health-related information that can be easily implemented [38,49]. Improving accessibility by providing health-related information in readable and plain language is outlined in principle 11. It is widely agreed that written health-related information should not be above an eighth-grade reading level [50,51], and for greater inclusivity, no higher than fifth grade has been suggested [52]. Generally, readability has not been included in widely used quality assessment tools, although it has often been evaluated alongside information quality in health research [27]. The inclusion of principle 11 in PRHISM is supported by literature that emphasizes the importance of providing written and nonwritten health-related information in simple and plain language [42,52], and recent research indicates that this is also pertinent to social media [53]. For the information to be accessible, it should be easily understood the first time it is heard or read [42].

Principle 12 specifies that health-related social media content should be accessible to individuals with vision and hearing impairments. Those living with vision and hearing impairments typically have poor health literacy and face challenges when accessing health-related information [54]. There have been recent calls to address this issue and provide guidance to health professionals to deliver accessible information [54]. Interestingly, principles 3, 11, and 12, which relate to various components of accessibility, were new principles suggested by the participants, and such considerations have not been included in previous tools [28-31,39]. This may reflect a greater understanding among experts regarding the importance of health-related information that meets the needs of all members of the population. The need for accessible health-related information has been advocated in the literature and supports the inclusion of principles related to accessibility in PRHISM [38,42,49-52,54]. If adhered to, principles 3, 11, and 12 may assist consumers with diverse needs and improve their overall accessibility to health advice provided on social media.

Of the 13 principles included in PRHISM, 7 (54%) pertained to transparency. Principles 1 and 5 state the need for authors to specify their credentials and qualifications and provide details about the original sources of information used to compile the social media content, respectively. The components of information quality outlined in principles 1 and 5 were covered in all the tools included in the Delphi surveys [28-31], as well as several other commonly used quality assessment tools [27], indicating their importance. Providing financial and conflict of interest disclosures is specified in principle 4 of PRHISM and received very strong support from participants (34/34, 100% rated as important or very important). As social media marketing and *influencing* has expanded, advertising associations have released statements declaring the need for prominent and clear disclosures of advertisements and other conflicts of interest in social media content [43,44]. The importance of providing such disclosures in health-related social media content has also been echoed in recent literature [53], and failure to disclose relevant conflicts of interest has the power to erode the public's trust in authoritative voices [55].

Transparency through the provision of comprehensive and balanced information is stipulated under principles 6 and 7 of PRHISM. These principles state that health-related claims should provide complete information on risks and benefits and clearly outline the limitations or areas of uncertainty. Comprehensive and balanced information provision is fundamental to the DISCERN instrument [28] and QUEST [39]. Furthermore, disclosing all relevant contexts and limitations in health-related social media content has been outlined as necessary to ensure that consumers can reach informed conclusions [53]. Others have acknowledged that although transparency in health communication is important, it is complex and, in some instances, may work against public understanding [56]. For example, research has shown that transparent information on scientific uncertainties and the risks and benefits of certain health behaviors has led to consumer confusion, which can undermine the public's trust in science [56,57]. Transparency is emphasized throughout all prominent quality assessment tools [27-31] and has been highlighted as critical to the credibility of health-related information sources on social media [53,58], supporting its centrality to most of the items included in PRHISM. This creates a challenge for content creators in balancing elements of transparency and accessibility to ensure that information is complete, credible, and understandable.

The need for quality health-related information on social media to be evidence based and authoritative is covered in principles 2 and 6 of PRHISM. Such considerations are not new and have long been recognized as fundamental to the provision of high-quality science and health-related information [27]. Recently, authoritative sources and evidence-based information have been identified as key credibility attributes for health-related social media content [53,58]. Others have highlighted the need for propagation of evidence-based information on social media to curtail the influence of misinformation on health issues such as vaccine hesitancy [59] and have also noted the responsibility of authoritative voices to combat and refute sources of misinformation on social media [60].

Finally, principles 9 and 10 are centered on supporting the relationships between patients and their health care providers. The DISCERN instrument, HONcode principles, and QUEST all express the importance of health-related information that is complementary to advice from health professionals [28,29,31,39]. Although social media can be a useful and powerful tool for the dissemination of health-related information [60], most information is not personalized and does not consider individuals' medical history or health needs. Codes of conduct apply to registered health care providers, and the provision of personalized health advice via social media is considered unethical and has resulted in the formal investigation of medical professionals [61,62]. Principles 9 and 10 accord with the health information–seeking behavior literature, which acknowledges that although web-based information environments can be interactive, they are not sufficient to replace the tailored care and advice that professionals can provide in a health care setting [6,63]. Research has shown that discussing web-based health-related information with a relevant professional can help strengthen patient-physician relationships and bolster shared decision-making [63,64].

### Strengths and Limitations

This study had a number of strengths. First, the Delphi technique was a highly appropriate method for meeting the aim of this research, and the sample size was suitable to produce meaningful results [35,40]. Second, the retention rate (69% retention of round 1 participants) was high, particularly given the generally low retention rates observed in Delphi studies [35,40]. Third, most of the principles included in PRHISM were adapted from tools that are widely used and have been shown to have good reliability [27-31,39]. Fourth, PRHISM is suitable for the evaluation of information from all health-related disciplines, and the accompanying scoring tool and guide for use will assist researchers when using the tool. Finally, PRHISM can also be used by health professionals to guide the development of quality health-related social media content and inform consumers about the attributes of quality information.

This investigation also had a number of limitations. First, academics, researchers, and respondents from Australia were overrepresented in the study sample, which may have limited the range of the perspectives captured. In addition, a general limitation of the Delphi method is that consensus does not necessarily mean that the correct answer has been chosen [35], which was evident in this study when 2 opposing principles remained after round 3. The research team conferred and discussed the face validity of these principles, informed by the literature and the overall Delphi process to mitigate the impact, to form a final decision. Fourth, principle 13 focused on image-based content and did not mention video content. This principle may have been more comprehensive if it also focused on video content; however, this was not considered until after the principles and PRHISM guide had been sent to participants for final comment. Finally, although PRHISM has been designed to be flexible to the changing social media landscape, the tool may require updates as social media continues to evolve.

### Implications and Future Directions

PRHISM may contribute to measuring and improving the quality of health-related information on social media. However, other strategies will also be required. Greater efforts by social media platforms to prevent the propagation of misinformation and direct consumers to credible information are needed. Some social media companies have taken steps to remove or provide warnings about health-related misinformation on their platforms [65,66]. Nevertheless, their impact is likely insufficient, and greater regulation, moderation, and fact checking by social media platforms are needed [53,59]. A practical implication of PRHISM is that the principles may be used to inform such regulation of social media platforms to make high-quality health information more prominent and limit the propagation of low-quality information. Greater efforts are also required to provide education to improve consumers' health and media literacy so that they are better able to identify credible health-related information from dangerous misinformation. A further practical implication of this study is that the principles may form the basis of educational materials to help develop health and media literacy. Finally, PRHISM is the first quality assessment tool for health information specific to social media and provides a standardized measure of information quality. Future studies that aim to assess the quality of health-related social media content should use PRHISM, instead of quality assessment tools that are not specific to social media, to improve the measurement of health information in social media research.

## Conclusions

Previously developed quality assessment tools for evaluating health-related information are not appropriate for evaluating social media content as they do not consider the unique characteristics of social media, and this study served to address this gap. Resulting from a comprehensive Delphi process, PRHISM comprised 13 principles that can be used by researchers to evaluate the quality of health-related information on social media. The principles promote accessibility, transparency, and authoritative and evidence-based information provision, supporting relationships between consumers and health care providers. The information contained in the PRHISM guide (Multimedia Appendix 2) defines what can be considered low-, moderate-, and high-quality information and sets out instructions for use in a research setting. Information from the PRHISM guide can also be used by health professionals and content creators to inform the provision of high-quality health-related social media content and by consumers to help identify high- and low-quality information. Further research is needed to improve the media and health literacy skills of the general population and regulate misleading poor-quality health-related information on social media platforms to reduce the threat of misinformation to public health.

## Acknowledgments

ED was supported by a Deakin University Postgraduate Research Scholarship. The authors would like to thank the experts who participated in this study.

## Authors' Contributions

ED contributed to the study design, data collection, data analysis, and development of the Principles for Health-related Information on Social Media tool and manuscript. SAM and RL supervised the research; contributed to the study design, content analysis, and interpretation of results; and assisted with the preparation of the Principles for Health-related Information on Social Media tool and manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables - Principles for Health-related Information on Social Media Delphi Round 2 and 3 consensus items and adaptations.

[DOCX File , 30 KB - jmir_v24i9e37337_app1.docx ]

Multimedia Appendix 2

Principles for Health-related Information on Social Media (PRHISM) guide and scoring tool.

[DOCX File , 108 KB - jmir_v24i9e37337_app2.docx ]

## References

1. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
2. 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]
3. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
4. Fox S. The social life of health information. Pew Research Center. 2014 Jan 15. URL: https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2019-12-21]
5. Alduraywish SA, Altamimi LA, Aldhuwayhi RA, AlZamil LR, Alzeghayer LY, Alsaleh FS, et al. Sources of health information and their impacts on medical knowledge perception among the Saudi Arabian population: cross-sectional study. J Med Internet Res 2020 Mar 19;22(3):e14414 [FREE Full text] [doi: 10.2196/14414] [Medline: 32191208]
6. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Social Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
7. Hirshberg P. First the media, then us: how the internet changed the fundamental nature of the communication and its relationship with the audience. OpenMind BBVA. URL: https://www.bbvaopenmind.com/en/articles/first-the-media-then-us-how-the-internet-changed-the-fundamental-nature-of-the-communication-and-its-relationship-with-the-audience/ [accessed 2021-02-13]
8. Fox S. The social life of health information, 2011. Pew Research Center. 2011 May 12. URL: https://www.pewresearch.org/internet/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2020-10-15]
9. Fassier P, Chhim A, Andreeva VA, Hercberg S, Latino-Martel P, Pouchieu C, et al. Seeking health- and nutrition-related information on the Internet in a large population of French adults: results of the NutriNet-Santé study. Br J Nutr 2016 Apr 15;115(11):2039-2046. [doi: 10.1017/s0007114516001355]
10. Obar JA, Wildman S. Social media definition and the governance challenge: an introduction to the special issue. Telecommunications Policy 2015 Oct;39(9):745-750 [FREE Full text] [doi: 10.1016/j.telpol.2015.07.014]
11. Dentzel D. How the internet has changed everyday life. Open Mind BBVA. URL: https://www.bbvaopenmind.com/en/articles/internet-changed-everyday-life/ [accessed 2021-02-13]
12. Kemp S. Digital 2021: global overview report. Data Reportal. 2021 Jan 27. URL: https://datareportal.com/reports/digital-2021-global-overview-report [accessed 2021-03-01]
13. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013 Jan 30;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]

14. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283 [FREE Full text] [doi: 10.1111/hir.12192] [Medline: 29045011]
15. Lambert M, Chivers P, Farringdon F. In their own words: a qualitative study exploring influences on the food choices of university students. Health Promot J Austr 2019 Jan 25;30(1):66-75. [doi: 10.1002/hpja.180] [Medline: 29892990]
16. Huo J, Desai R, Hong Y, Turner K, Mainous AG, Bian J. Use of social media in health communication: findings from the health information national trends survey 2013, 2014, and 2017. Cancer Control 2019 Apr 18;26(1):1073274819841442 [FREE Full text] [doi: 10.1177/1073274819841442] [Medline: 30995864]
17. Zucker HA. Tackling online misinformation: a critical component of effective public health response in the 21st century. Am J Public Health 2020 Oct;110(S3):S269. [doi: 10.2105/AJPH.2020.305942] [Medline: 33001725]
18. Merchant RM, South EC, Lurie N. Public health messaging in an era of social media. JAMA 2021 Jan 19;325(3):223-224. [doi: 10.1001/jama.2020.24514] [Medline: 33393964]
19. Global Risks 2013 Eighth Edition. World Economic Forum. URL: https://reports.weforum.org/global-risks-2013/ [accessed 2019-12-01]
20. Chen Y, Conroy NK, Rubin VL. News in an online world: the need for an “automatic crap detector”. Proc Assoc Info Sci Tech 2016 Feb 24;52(1):1-4. [doi: 10.1002/pra2.2015.145052010081]
21. Yates B. Media literacy: a health education perspective. J Health Educ 2013 Feb 22;30(3):180-184 [FREE Full text] [doi: 10.1080/10556699.1999.10603399]
22. Naeem SB, Bhatti R, Khan A. An exploration of how fake news is taking over social media and putting public health at risk. Health Info Libr J 2021 Jun;38(2):143-149 [FREE Full text] [doi: 10.1111/hir.12320] [Medline: 32657000]
23. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
24. Hussain A, Ali S, Ahmed M, Hussain S. The anti-vaccination movement: a regression in modern medicine. Cureus 2018 Jul 03;10(7):e2919 [FREE Full text] [doi: 10.7759/cureus.2919] [Medline: 30186724]
25. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
26. Afful-Dadzie E, Afful-Dadzie A, Egala SB. Social media in health communication: a literature review of information quality. Health Inf Manag 2021 Apr 04:1833358321992683. [doi: 10.1177/1833358321992683] [Medline: 33818176]
27. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
28. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
29. Charnock D. The DISCERN Handbook. Abingdon: Radcliffe Medical Press; 1998.
30. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet: caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
31. Barrett S. The Significance of the HONcode principles. Quackwatch. 2021 Feb 01. URL: https://quackwatch.org/about/honcode/ [accessed 2022-08-30]
32. Dalmer NK. Questioning reliability assessments of health information on social media. J Med Libr Assoc 2017 Jan 17;105(1):61-68 [FREE Full text] [doi: 10.5195/jmla.2017.108] [Medline: 28096748]
33. Geyser W. What is the best length for social media posts? Influencer Marketing Hub. 2022 Aug 16. URL: https://influencermarketinghub.com/best-length-for-social-media-posts/ [accessed 2021-07-07]
34. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]
35. Trevelyan E, Robinson P. Delphi methodology in health research: how to do it? Eur J Integrative Med 2015 Aug;7(4):423-428 [FREE Full text] [doi: 10.1016/j.eujim.2015.07.002]
36. Stewart D, Gibson-Smith K, MacLure K, Mair A, Alonso A, Codina C, et al. A modified Delphi study to determine the level of consensus across the European Union on the structures, processes and desired outcomes of the management of polypharmacy in older people. PLoS One 2017 Nov 20;12(11):e0188348 [FREE Full text] [doi: 10.1371/journal.pone.0188348] [Medline: 29155870]
37. Vogel C, Zwolinsky S, Griffiths C, Hobbs M, Henderson E, Wilkins E. A Delphi study to build consensus on the definition and use of big data in obesity research. Int J Obes (Lond) 2019 Dec;43(12):2573-2586 [FREE Full text] [doi: 10.1038/s41366-018-0313-9] [Medline: 30655580]
38. Mete R, Kellett J, Bacon R, Shield A, Murray K. The P.O.S.T guidelines for nutrition blogs: a modified e-Delphi study. J Acad Nutr Diet 2021 Jul;121(7):1219-1230 [FREE Full text] [doi: 10.1016/j.jand.2021.02.014] [Medline: 33781728]

39. Robillard JM, Jun JH, Lai J, Feng TL. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inform Decis Mak 2018 Oct 19;18(1):87 [FREE Full text] [doi: 10.1186/s12911-018-0668-9] [Medline: 30340488]
40. Keeney S, Hasson F, McKenna H. The Delphi Technique in Nursing and Health Research. West Sussex: Wiley-Blackwell; 2010.
41. Holey EA, Feeley JL, Dixon J, Whittaker VJ. An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. BMC Med Res Methodol 2007 Nov 29;7(1):52 [FREE Full text] [doi: 10.1186/1471-2288-7-52] [Medline: 18045508]
42. Wicklund K, Ramos K. Plain language: effective communication in the health care setting. J Hospital Librarianship 2009 Apr 22;9(2):177-185. [doi: 10.1080/15323260902812740]
43. Clearly distinguishable advertising. Australian Association of National Advertisers. 2016. URL: https://aana.com.au/content/uploads/2017/01/AANA_Distinguishable-Advertising-Best-Practice-Guideline__Final.pdf [accessed 2021-02-14]
44. Australian influencer marketing code of practice. Australian Influencer Marketing Council. 2020 Jul 1. URL: https://www.auditedmedia.org.au/site/assets/media/AIMCO-docs/AIMCO-Code-of-Practice-July-1-2020-updated-6-Oct.pdf [accessed 2021-02-14]
45. Sharma R. Why you should add Instagram Alt Text to your posts (+ Tutorial!). Tailwind. 2021 Apr 1. URL: https://www.tailwindapp.com/blog/instagram-alt-text [accessed 2021-06-07]
46. Leetaru K. Facebook's terrible automatic video captions are no laughing matter. Forbes. 2019 Apr 19. URL: https://www.forbes.com/sites/kalevleetaru/2019/04/19/facebooks-terrible-automatic-video-captions-are-no-laughing-matter/?sh=11be3a87798f [accessed 2021-06-05]
47. Social media accessibility guidelines. Princeton University. URL: https://accessibility.princeton.edu/guidelines/social-media [accessed 2021-06-05]
48. Roselli A. Improving your tweet accessibility. Adrian Roselli. URL: https://adrianroselli.com/2018/01/improving-your-tweet-accessibility.html#Captions [accessed 2021-06-05]
49. Worsley A. Nutrition knowledge and food consumption: can nutrition knowledge change food behaviour? Asia Pac J Clin Nutr 2002;11 Suppl 3(s3):S579-S585 [FREE Full text] [doi: 10.1046/j.1440-6047.11.supp3.7.x] [Medline: 12492651]
50. Oliffe M, Thompson E, Johnston J, Freeman D, Bagga H, Wong PK. Assessing the readability and patient comprehension of rheumatology medicine information sheets: a cross-sectional Health Literacy Study. BMJ Open 2019 Feb 05;9(2):e024582 [FREE Full text] [doi: 10.1136/bmjopen-2018-024582] [Medline: 30813117]
51. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. Clin Orthop Relat Res 2010 Oct;468(10):2572-2580 [FREE Full text] [doi: 10.1007/s11999-010-1380-y] [Medline: 20496023]
52. Communicating with patients who have limited literacy skills. Report of the National work group on literacy and health. J Fam Pract 1998 Feb;46(2):168-176. [Medline: 9487325]
53. Kington RS, Arnesen S, Chou WS, Curry SJ, Lazer D, Villarruel AM. Identifying credible sources of health information in social media: principles and attributes. NAM Perspect 2021;2021:10.31478/202107a [FREE Full text] [doi: 10.31478/202107a] [Medline: 34611600]
54. Withers J, Speight C. Health care for individuals with hearing loss or vision loss: a minefield of barriers to accessibility. N C Med J 2017 Apr 18;78(2):107-112 [FREE Full text] [doi: 10.18043/ncm.78.2.107] [Medline: 28420772]
55. Garza C, Stover PJ, Ohlhorst SD, Field MS, Steinbrook R, Rowe S, et al. Best practices in nutrition science to earn and keep the public's trust. Am J Clin Nutr 2019 Jan 01;109(1):225-243 [FREE Full text] [doi: 10.1093/ajcn/nqy337] [Medline: 30657846]
56. Rowe S, Alexander N. Nutrition science communication. Nutr Today 2021;56(2):76-79. [doi: 10.1097/nt.0000000000000468]
57. Holmberg C. Politicization of the low-carb high-fat diet in Sweden, promoted on social media by non-conventional experts. Int J E Politc 2015;6(3). [doi: 10.4018/ijep.2015070103]
58. WHO online consultation meeting to discuss global principles for identifying credible sources of health information on social media. World Health Organization. 2022 Jul 1. URL: https://www.who.int/publications/m/item/who-online-consultation-meeting-to-discuss-global-principles-for-identifying-credible-sources-of-health-information-on-social-media [accessed 2022-07-01]
59. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
60. Szabo RA. Hippocrates would be on Twitter. Med J Aust 2020 Dec 23;213(11):506-7.e1. [doi: 10.5694/mja2.50873] [Medline: 33230809]
61. Noakes TD. Use of social media by health professionals in South Africa. S Afr Med J 2017 Aug 25;107(9):724. [doi: 10.7196/samj.2017.v107i9.12709]
62. Social media: how to meet your obligations under the National Law. Medical Board AHPRA. 2019 Nov. URL: https://www.medicalboard.gov.au/Codes-Guidelines-Policies/Social-media-guidance.aspx [accessed 2021-12-15]
63. Wald HS, Dube CE, Anthony DC. Untangling the Web--the impact of Internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]

64. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
65. COVID-19 misleading information policy. Twitter. 2021 Dec. URL: https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy [accessed 2021-12-02]
66. Keeping people informed, safe, and supported on Instagram. Instagram. URL: https://about.instagram.com/blog/announcements/coronavirus-keeping-people-safe-informed-and-supported-on-instagram [accessed 2021-12-02]

## Abbreviations

**HONcode:** Health on the Net
**JAMA:** Journal of the American Medical Association
**PRHISM:** Principles for Health-related Information on Social Media
**QUEST:** Quality Evaluation Scoring Tool

*Edited by A Mavragani; submitted 16.02.22; peer-reviewed by J Marquard, A Pal; comments to author 07.04.22; revised version received 01.06.22; accepted 27.06.22; published 08.09.22.*

*Please cite as:*
*Denniss E, Lindberg R, McNaughton SA*
*Development of Principles for Health-Related Information on Social Media: Delphi Study*
*J Med Internet Res 2022;24(9):e37337*
*URL: https://www.jmir.org/2022/9/e37337*
*doi:10.2196/37337*
*PMID:36074544*

Original Paper

# Psychosocial Needs of Gynecological Cancer Survivors: Mixed Methods Study

Elizabeth J Adams[1], BA; David Tallman[2], BS; Marcy L Haynam[2], MS; Larissa Nekhlyudov[3], MD, MPH; Maryam B Lustberg[4], MD, MPH

[1]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[2]Wexner Medical Center, The Ohio State University, Columbus, OH, United States
[3]Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, United States
[4]Yale Cancer Center, School of Medicine, Yale University, New Haven, CT, United States

**Corresponding Author:**
Elizabeth J Adams, BA
Feinberg School of Medicine
Northwestern University
420 E Superior St
Chicago, IL, 60611
United States
Phone: 1 312 503 8649
Email: elizabeth.adams@northwestern.edu

## Abstract

**Background:** Internet and social media platforms offer insights into the lived experiences of survivors of cancer and their caregivers; however, the volume of narrative data available is often cumbersome for thorough analysis. Survivors of gynecological cancer have unique needs, such as those related to a genetic predisposition to future cancers, impact of cancer on sexual health, the advanced stage at which many are diagnosed, and the influx of new therapeutic approaches.

**Objective:** This study aimed to present a unique methodology to leverage large amounts of data from internet-based platforms for mixed methods analysis. We analyzed discussion board posts made by survivors of gynecological cancer on the American Cancer Society website with a particular interest in evaluating the psychosocial aspects of survivorship.

**Methods:** All posts from the ovarian, uterine, and gynecological cancers (other than ovarian and uterine) discussion boards on the American Cancer Society Cancer Survivors Network were included. Posts were web scraped using Python and organized by psychosocial themes described in the Quality of Cancer Survivorship Care Framework. Keywords related to each theme were generated and verified. Keywords identified posts related to the predetermined psychosocial themes. Quantitative analysis was completed using Python and R Foundation for Statistical Computing packages. Qualitative analysis was completed on a subset of posts as a proof of concept. Themes discovered through latent Dirichlet allocation (LDA), an unsupervised topic modeling technique, were assessed and compared with the predetermined themes of interest.

**Results:** A total of 125,498 posts made by 6436 survivors of gynecological cancer and caregivers between July 2000 and February 2020 were evaluated. Of the 125,489 posts, 23,458 (18.69%) were related to the psychosocial experience of cancer and were included in the mixed methods psychosocial analysis. Quantitative analysis (23,458 posts) revealed that survivors across all gynecological cancer discussion boards most frequently discussed the role of friends and family in care, as well as fatigue, the effect of cancer on interpersonal relationships, and health insurance status. Words related to psychosocial aspects of survivorship most often used in posts included “family,” “hope,” and “help.” Qualitative analysis (20 of the 23,458 posts) similarly demonstrated that survivors frequently discussed coping strategies, distress and worry, the role of family and caregivers in their cancer care, and the toll of managing financial and insurance concerns. Using LDA, we discovered 8 themes, none of which were directly related to psychosocial aspects of survivorship. Of the 56 keywords identified by LDA, 2 (4%), “sleep” and “work,” were included in the keyword list that we independently devised.

**Conclusions:** Web-based discussion platforms offer a great opportunity to learn about patient experiences of survivorship. Our novel methodology expedites the quantitative and qualitative analyses of such robust data, which may be used for additional patient populations.

*(J Med Internet Res 2022;24(9):e37757)* doi:10.2196/37757

**KEYWORDS**

mixed methods; quantitative; qualitative; cancer survivorship; gynecological cancer; uterine cancer; ovarian cancer; cancer informatics; patient discussion; social media

## Introduction

### Background

Gynecological cancer is a broad disease category that includes cervical, uterine, ovarian, vaginal, and vulvar cancers, which are distinct in their presentation, pathology, treatment, prognosis, and survivorship trajectories. It is estimated that there were approximately 116,760 new cases and 34,080 new deaths from gynecological cancers in 2021 [1]. The 5-year overall survival rates for uterine and ovarian cancer are now 81% and 48%, respectively [1]. Although there is vast room for improvement in these survival rates through earlier detection and improved treatment, there has been notable progress in the management and survival of those with gynecological cancer over time. Thus, there is a larger population of survivors of gynecological cancer now than ever before. With this growing population of survivors of gynecological cancer, there is a paramount need to investigate methods to both assess and implement survivorship-focused care among survivors diagnosed and living with gynecological cancer.

Cancer survivorship was first described in 2006 in the landmark publication "From Cancer Patient to Cancer Survivor: Lost in Transition" [2], which outlined the gap in care between active cancer treatment and life after completion of active cancer treatment. Much work has been conducted in survivorship since 2006, with an emphasis on survivorship beginning on the day of the cancer diagnosis. This is reflected in the National Cancer Institute's definition of survivorship: the "health and well-being of a person with cancer from the time of diagnosis until the end of life" [3]. Thus, survivorship care is focused on every aspect of health and wellness that is not directly related to treating the cancer itself.

The National Comprehensive Cancer Network provides guidelines on the delivery of survivorship care and outlines 7 aspects of survivorship care, including preventing new and recurrent cancers, preventing late effects of cancer and treatment, surveillance for return of cancer, screening for new cancers, assessing and treating late effects of cancer and treatment, coordinating care among providers, and planning for ongoing survivorship care [4]. Additional guidelines, such as those by the American Cancer Society (ACS), the American Society of Clinical Oncology, and Cancer Care Ontario, assert similar components as essential aspects of survivorship care [5-7]. The Quality of Cancer Survivorship Care Framework defines domains of survivorship care through an iterative review of survivorship care guidelines that was bolstered with key expert interviews [8]. All guidelines emphasize the importance of considering the psychosocial aspects of survivorship, which the National Cancer Institute defines as the "mental, emotional, social, and spiritual effects of a disease" with the psychosocial effects of cancer, including "changes in how a patient thinks, their feelings, moods, beliefs, ways of coping, and relationships with family, friends, and co-workers" [9]. The psychosocial effects of cancer and cancer treatment are an important area for exploration as they are unique to each cancer type, cancer treatment, and individual patient.

### Digitalizing Survivorship

As survivorship is a relatively new component of oncologic care, the field continues to develop new interventions and modify existing initiatives to best meet the needs of survivors. Learning directly from the voices of survivors is an essential step to inform these survivorship services. Web-based discussion boards and social media platforms have become tools for researchers and clinicians to harness publicly available discussions on cancer survivorship, offering extremely large quantities of candid and spontaneous thoughts and opinions from survivors [10]. Gao et al [11] recently demonstrated the utility of investigating posts made on Instagram related to head and neck cancer, where they reported that most of the posts made by patients were focused on appointments, treatment, and side effects. Similar analyses have been conducted on Twitter [12] and Facebook [13] posts for additional cancer types. Discussion-based forums, including Facebook groups [14] and those on the ACS website [15], have also been analyzed by researchers to acquire a better understanding of survivorship from the perspective of survivors of cancer. Although these prior studies demonstrated the utility of analyzing posts and discussions among survivors, each was conducted with manual thematic analysis. The feasibility of organizing and analyzing text-based posts and discussions by hand has an upper limit, which undermines one of the greatest strengths of exploring such posts—the volume of data available. To overcome this, other researchers have used automated technology such as unsupervised topic modeling to determine the most prevalent themes in web-based and social media posts [16-18]. However, this process does not necessarily identify themes that are of importance to the researcher, such as when the themes are related to particular aspects of survivorship care, and does not detect predetermined themes of interest.

### Objective

The ACS Cancer Survivors Network has discussion boards specific to types of cancer, where survivors and caregivers can interact through posts. To demonstrate a novel methodology to scan through an extensive compilation of these posts and thematically organize them, we analyzed the ACS gynecological cancer discussion boards, including the ovarian, uterine, and gynecological cancers (other than ovarian and uterine) discussion boards. Our methodology automated the analysis of >125,000 discussion board posts using the previously described Quality of Cancer Survivorship Care Framework, with a specific focus on the themes within the framework's *Surveillance and Management of Psychosocial Effects Domain* [8]. Our goal was to demonstrate the utility of this novel methodology to expedite the guided quantitative and qualitative analysis of a robust amount of discussion-based data from survivors of gynecological cancer and caregivers based on predetermined themes of interest, such as those in the Quality of Cancer Survivorship Care Framework.

## *Methods*

### Study Design

This study uses a mixed methods approach, specifically a sequential explanatory design [19]. Sequential explanatory studies involve two phases in the analysis. First, a quantitative analysis is conducted, which informs a second, qualitative analysis phase [19,20]. The information from the first quantitative analysis and second qualitative analysis is combined using integration strategies. Per the standard sequential explanatory study design, there are 2 stages to data integration, with the first taking place after the conclusion of quantitative analysis. This was completed by determining which themes were most prevalent in the data set, which allowed us to connect the quantitative component of the study to the qualitative component. The second stage of data integration took place after the completion of qualitative analysis, during which we compared the quantitative and qualitative findings to share the reported findings [19,20]. Viewing both the findings from the quantitative analysis and those from the qualitative analysis as contributing equally to the results but in different ways is referred to as complementary stance integration [21]. To implement this, we completed the following iterative process that will be detailed in the paragraphs that follow: identify predetermined themes of interest, web scrape, develop keywords to capture themes in the web-based text, verify the validity and reliability of the keywords, apply the keywords to the data set, and analyze the data output.

This process was developed after an initial exploration of topic modeling, specifically latent Dirichlet allocation (LDA). Topic modeling is an unsupervised machine learning approach to determine patterns of related words in large quantities of text, thus independently discovering themes that the program determines to be of significance based on probability [22]. The Python packages *gensim* and *LDAvis* were used to facilitate theme discovery. To determine the best number of topics for our data set, the metric of coherence ($C_V$) was used [23]. Models of varying numbers of topics ranging from 2 to 40 were developed in increments of 2, and the highest $C_V$ score of 0.52 was observed with 8 themes. A relevance metric in the LDAvis package was used to evaluate the most relevant keywords rather than just the most prevalent as there was a high frequency of particular words (eg, “cancer”) based on the nature of the discussion board data. On the basis of correlation with human interpretability, the value of $\lambda$ was set to 0.6 [24]. Consensus was reached among the study team on the semantic meaning of the 8 topics. These topic modeling–generated themes were compared with those in the Quality of Cancer Survivorship Care Framework [8]. A second attempt at LDA was used to generate a model with 16 themes, which had the second highest $C_V$ score, as a means of more closely mimicking the number of psychosocial themes in the Quality of Cancer Survivorship Care Framework to obtain topics with more narrow themes.

### Predetermined Themes of Interest

The Quality of Cancer Survivorship Care Framework [8] has multiple quality domains, including the *Surveillance and Management of Psychosocial Effects Domain*; *Surveillance, and Management of Physical EffectsDomain*; and *Care Coordination Domain*, among others. In this analysis, we focused on the *Surveillance and Management of Psychosocial Effects Domain*. Each domain contains several proposed quality indicators to assess different aspects of survivorship care; for example, within the *Surveillance and Management of Psychosocial Effects Domain*, the proposed quality indicators to assess cancer survivorship care include “fatigue” and “stress,” as well as “financial toxicity” and “fertility.” Although these indicators may be used to assess the quality of survivorship care, we felt that they encapsulated all of the relevant themes of interest that survivors of gynecological cancer may discuss on the web on the ACS discussion boards. Therefore, for the purposes of our study, we will be referring to the indicators listed in the framework’s *Surveillance and Management of Psychosocial Effects Domain* as “themes.” We use these themes to demonstrate the utility of our novel methodology by detecting the presence of these themes in communications between survivors of gynecological cancer and caregivers.

### Web Scraping

All 125,498 posts from the ovarian, uterine, and gynecological cancers (other than ovarian and uterine) ACS discussion boards were evaluated. The posts included in the analysis were created from July 21, 2000, through February 24, 2020, the date when web scraping was completed. Posts were either responses that were added to an existing conversation or the “conversation-starters” on the discussion board. Web scraping, or simply “scraping,” is a technique used to extract the content of interest from web-based platforms so that it can be analyzed using computer software, essentially “downloading” it in a way that can be used by researchers. Python is a computer programming language that automates specific actions performed on the computer, such as the process of web scraping. By creating a custom Python script, the process of scraping >125,000 posts was automated. Python has multiple packages that allow for the software to perform different actions. In our analysis, we specifically used the Python packages *urllib* and *Beautiful Soup* 4 to navigate and extract the text from discussion posts on the ACS discussion boards. First, *urllib* was used to interface with the ACS website and scrape the web page. Next, *Beautiful Soup* 4 was used to parse the HTML code obtained from the ACS web pages. Together, this allowed us to extract the submitted text contained in a post, the date of the post, and the username of the poster. After the posts were scraped from the ovarian, uterine, and gynecological cancers (other than ovarian and uterine) discussion boards, they were saved to a CSV data file that was used for further downstream analyses.

### Capturing Themes of Interest Using Keywords

We devised a list of “keywords” that captured each of the predetermined themes from the Quality of Cancer Survivorship

Care Framework's *Surveillance and Management of Psychosocial Effects Domain* (Textbox 1). Every theme had multiple keywords that were created from synonyms, related phrases, and related words. The keyword list used truncated, root words to capture all variations of a given keyword; for example, 1 theme in the *Surveillance and Management of Psychosocial Effects Domain* is "underemployment, unemployment, return to work." One of the numerous keywords to capture this theme is based on the root word "unemploy." By using the root word "unemploy," the methodology captures all variations of it, including "unemployment," "unemployed," and "unemployable." This allows us to capture all instances when survivors may have discussed topics related to the base word "unemploy."

The purpose of the keyword list was to be able to determine which ACS posts discussed any of the predetermined themes of interest. Our methodology scans through data, detecting when the keywords were used. The use of a particular theme's keywords indicated to us that the ACS post discussed that particular theme. The software noted how many times the keywords were present from every single predetermined theme across every single evaluated ACS post. Thus, each ACS post was assigned a "theme score" for every theme. This allowed us to determine which themes were most prevalent in a given post without reading it first based on the number of times that theme's keywords appeared in it. Simultaneously, the theme score allowed us to immediately identify which ACS posts discussed specific themes of interest. This expedited the process as we were able to quickly pull up all of the posts relevant to particular themes, thus expediting additional review and qualitative analysis.

**Textbox 1.** Keyword list.

**Surveillance and Management of Psychosocial Effects themes and keywords**

**Psychological**

“Psychological” and “psychology”

**Fatigue**

“Fatigue,” “tired,” “tiring,” “fatiguing,” “exhaust,” “nap,” and “rest”

**Stress**

“Stress” and “stressed”

**Posttraumatic stress**

“Posttraumatic stress,” “trauma,” and “traumatized”

**Posttraumatic growth**

“Posttraumatic growth,” “trauma,” and “traumatized”

**Distress**

“Distress,” “depression,” “depressed,” “feel down,” “feeling down,” “sad,” “sadness,” “tear,” “cry,” “upset,” “heartbroken,” “heartbreaking,” “wrench,” “guilt,” and “cried”

**Anxiety**

“Anxiety,” “anxious,” and “panic”

**Fear of recurrences**

“Fear of recurrence” and “recurrence”

**Sleep disturbances**

“Sleep,” “insomnia,” “wake,” “asleep,” “disturb,” “restless,” and “sleep disturbance”

**Coping**

“Coping,” “cope,” and “coped”

**Worry**

“Worry,” “worried,” “worrying,” and “worries”

**Illness intrusiveness**

“lifestyle,” “intruding,” “illness intrusiveness,” “interfere,” “embarrass,” “ashamed,” “shame,” and “disrupt”

**Cognitive changes**

“Fog,” “memory,” “concentrate,” “concentrating,” “concentration,” “cognitive,” “cognition,” and “foggy”

**Educational problems**

“Educational problem,” “student,” “learning difficult,” and “problems in school”

**Social withdrawal**

“Social withdraw,” “withdraw socially,” “social isolation,” “socially isolating,” “lonely,” and “social withdrawal”

**Financial and employment**

“Financial,” “finances,” “employ,” “job,” “fulltime,” “full-time,” “part-time,” and “workload”

**Financial toxicity**

“Financial toxicity,” “debt,” “cost,” “bill,” “expensive,” “expense,” “money,” “money trouble,” and “financial trouble”

**Underemployment, unemployment, and return to work**

“Underemploy,” “unemploy,” “return to work,” “return fulltime,” “return full-time,” “return part-time,” “laid off,” “lay off,” “fire,” “quit,” and “fired”

**Work productivity**

“Work productivity,” “productive at work,” “working hard,” “work hard,” “falling behind,” “fall behind,” and “fell behind”

**School productivity**

“School productivity,” “learning,” and “school college”

**Insurance status**

"Insurance," "insured," "Medicaid," "Medicare," and "copay"

**Interpersonal**

"Interpersonal," "boyfriend," "husband," "spouse," "girlfriend," "wife," "significant other," "fiancé," "partner," and "relationship"

**Sexuality and intimacy**

"Sex," "intimacy," "intimate," "intercourse," "sexuality," and "sexual"

**Fertility**

"Fertility," "fertile," "infertility," "infertile," "preservation," "pregnancy," "pregnant," "conceiving," "conceive," "miscarriage," "miscarry," "IVF," "in vitro," "oocyte," "embryo," "freeze," "froze," "egg," "sperm," and "frozen"

**Family and caregiver relationships**

"Mother," "mom," "father," "dad," "sister," "brother," "son," "daughter," "friend," "spouse," "husband," "wife," "partner," "kid," "child," "family," "caregiver," "relationship," "friendship," "partnership," "marriage," "divorce," "separate," "engage," and "fiancé"

**Recommended evaluation provided (eg, laboratory testing, imaging, or referral to specialty care)**

"Psychological evaluation," "social history," "referral to a therapist," "referral to a psych," "referral to psych," "social work referral," "referral to social work," "referred to social work," "referred to a social work," and "referred to a psych"

**Treatment provided (eg, medication, therapy, or exercise)**

"Psychological treatment," "psychological medication," "counseling," "therapy," "support group," "Zoloft," "Xanax," "Lexapro," "Celexa," "Wellbutrin," "Desyrel," "Prozac," "Adderall," "Ativan," "Cymbalta," "Effexor," "Seroquel," and "Depakote"

**Assessment of adherence to treatment completed**

"Adherence," "adhere," "as instructed," "stick to," and "stuck with"

**Reassessment of symptoms and conditions at defined intervals or treatment phases**

"Reassessment of psychological symptoms," "reassess psychological symptoms," "review psychological symptoms," and "reviewed psychological symptoms"

## Keyword Verification

To determine whether our keywords were effective, a 2-step verification process was completed on 20 randomly selected posts from the gynecological cancer discussion boards. The goal of the verification process was to (1) check that the computer program appropriately categorized the keywords into their intended themes of interest based on the number of keywords present in the post and (2) verify that an individual naïve to the keywords would categorize the posts as being most related to the same theme that the computer program determined based on which theme had the greatest number of keywords represented. For the first step of keyword verification, a research team member (EA) manually assigned the 20 randomly selected posts to their psychosocial themes using the keyword list and noted which theme had the greatest number of keywords present in the post. EA created the keyword list and was therefore familiar with the keywords. This step was used to verify that the computer program properly captured each theme in a given post based on the number of keywords present for it.

The second step of the keyword verification process was completed by another research team member (MH), who had never seen the keyword list before. MH blindly assigned each of the posts to a theme within the *Surveillance and Management of Psychosocial Effects Domain* without the use of keywords. This step was used to verify that the keywords EA had created for each of the themes within the *Surveillance and Management of Psychosocial Effects Domain* were properly captured.

To determine whether the responses of EA and MH aligned with the computer program's responses for categorizing the 20 posts into their most relevant theme (the theme with the highest theme score), another research team member (DT) compared the 20 posts' theme designations by EA, MH, and the computer program. This keyword verification process was successful with a few minor discrepancies that were deemed acceptable as some of the predetermined themes were very closely related (eg, a survivor may be concerned that their "work productivity" could lead to them losing their job and becoming "unemployed," which are 2 different but strongly related themes). In the few occurrences of a discrepancy between the program and the researchers, the program still scored the posts very highly in the themes the researchers chose, meaning that the post still largely reflected the theme selected by the researchers and the computer program. Thus, a query to find ACS posts discussing themes designated by either the researchers or the computer program would have highlighted the post.

## Applying the Keywords

Once we were confident that our keywords and methodology captured the themes from the *Surveillance and Management of Psychosocial Effects Domain*, we used the methodology at large for all 125,489 posts. To automate the generation of the "theme score" for all posts and all of the predetermined themes of interest, we created another Python script using the base string library. Each post had a theme score calculated for every theme, with the theme score reflecting the number of times the theme's keywords appeared in the post.

Next, we were able to sort the posts' theme scores for every theme. This was done so that we could find the most relevant posts within a given theme. Simultaneously, we could see which themes were most prevalent in a given post. For us to consider a post to be "related" to a particular theme, we set a minimum theme score of 3 (a post must contain ≥3 instances of a theme's keywords). Setting a minimum theme score minimized the number of times someone used a keyword spuriously in a way that was not necessarily relevant to the predetermined theme of interest; for example, a survivor writing "unemployed" once in their post would not prompt the computer program to label it as being related to the "Underemployment, unemployment, return to work" theme.

### Ethics Approval and Data Use

Publicly available data, such as the posts contained on the ACS discussion boards, do not constitute research with human subjects per the Office for Human Research Protections. Thus, institutional review board review approval was not required for the conduction of data analysis, interpretation, and dissemination of findings, as supported by 45CFR46:102 [25]. This is in line with the institutional review board process of all affiliated institutions involved in this analysis.

### Analysis

#### *Quantitative Analysis*

The process of web scraping and data collection using our methodology generated various quantitative data points to analyze and describe graphically using R and the *ggplot2* package for plotting; for example, we determined which predetermined themes of interest were most frequently discussed on the ovarian, uterine, and gynecological cancers (other than ovarian and uterine) discussion boards. We also examined trends over time across the discussion boards in terms of discussed themes. We determined the prevalence of each theme by taking the number of posts related to a particular theme and dividing it by the total number of posts that year.

Predetermined themes of interest and their respective keywords aside, we wished to pictorially display the words that survivors and caregivers used in their posts. To capture this, we created a graph depicting the 40 most frequently used words in the ACS posts from all 125,498 posts.

#### *Qualitative Analysis*

By using the keyword list and capturing the most salient predetermined themes in a given post through its theme score, we could easily determine which ACS posts were related to particular themes of interest. This facilitated further qualitative analysis; for example, if we wished to see *how* survivors and caregivers spoke about the theme "Underemployment, unemployment, return to work," then we could sort through all 125,498 ACS posts to determine which had the greatest theme scores for "Underemployment, unemployment, return to work." Thus, this would streamline qualitative review such that the select posts related to the theme of interest were already identified and could promptly be analyzed.

To demonstrate how our methodology expedites qualitative analysis in a very large discussion-based data set, we qualitatively analyzed 20 posts. A sample size of 20 was selected based on previously described qualitative research recommendations [26,27]. We selected these 20 posts as they had the absolute greatest number of keyword instances across all of the predetermined themes. Stated another way, for each ACS post, we added up all of the "theme scores" (number of times the keywords were present for a given theme) from all of the themes. The 20 posts that were qualitatively reviewed had the greatest total when adding together all of the individual theme scores. As a number of ACS posters were quite active on the discussion boards, a maximum of 5 posts per user was included in the analysis. If the sixth post from a poster qualified for inclusion based on its theme score, it was instead excluded, and the next post from another unique user was included.

These qualitative analyses were completed by 2 research team members (EA and MH). Each post was individually examined, and the reviewers noted quotes that they deemed to be the most relevant to the themes in the *Surveillance and Management of Psychosocial Effects Domain*. Any discrepancies between the reviewers (EA and MH) that could not be resolved between them were resolved by a third reviewer (ML) to determine the final assigned themes. EA and MH also collected any demographic information available within the ACS posts about the survivor or caregiver who posted.

## *Results*

### Quantitative Analysis

A total of 125,498 posts were analyzed quantitatively, with 61,699 (49.16%) posts from the uterine cancer discussion board; 57,011 (45.43%) from the ovarian cancer discussion board; and 6788 (5.41%) from the gynecological cancers (other than ovarian and uterine) discussion board. These posts were created by a total of 6436 unique posters, with each unique poster creating an average of 19.5 (SD. 107.4, range 1-2397) posts over the 20-year period that was evaluated. The total number of posts experienced a large boost in 2008 and reached its maximum in 2011, with the ovarian cancer discussion board being the most prolific. The uterine cancer discussion board experienced a notable rise in posts in 2011 and 2016, and the gynecological cancers (other than ovarian and uterine) discussion board had a steady number of posts over time, as depicted in Figure 1. After this point, the number of posts on the ovarian cancer discussion board markedly decreased, whereas the uterine cancer discussion board maintained its higher number of posts and remained on an upward trend after this point.

Of the 125,489 total posts, 23,458 posts (18.69%) were related to the psychosocial experience of cancer and were further investigated based on the presence of at least a theme score of 3 from the *Surveillance and Management of Psychosocial Effects Domain*. As illustrated in Figure 2, the themes from the *Surveillance and Management of Psychosocial Effects Domain* had become more prevalent within the uterine cancer discussion board over time. The prevalence of these psychosocial themes of interest within the ovarian cancer and gynecological cancers (other than ovarian and uterine) discussion boards had decreased over time; however, this may be due in part to the decreasing number of overall posts within these 2 discussion boards. The most frequently discussed predetermined themes that survivors posted about across all of the gynecological cancer discussion boards were related to (1) the role of friends and family in care,

(2) fatigue, (3) the impact of interpersonal relationships, and (4) health insurance status. Additional prevalent themes of interest that were noted were related to stress, sexuality, and fertility.

The most prevalent words seen in the posts from all evaluated discussion boards, regardless of their relatedness to the predetermined psychosocial themes, are depicted in a graph (Figure 3). Most words used by survivors and caregivers in the ACS posts pertained to the treatment of cancer, including "cancer," "chemo," "treatment," "surgery," and "radiation." Many words were used by survivors that were related to survivorship based on psychosocial themes, including "family," "hope," "help," and "pain."

**Figure 1.** Number of posts by discussion board over time.

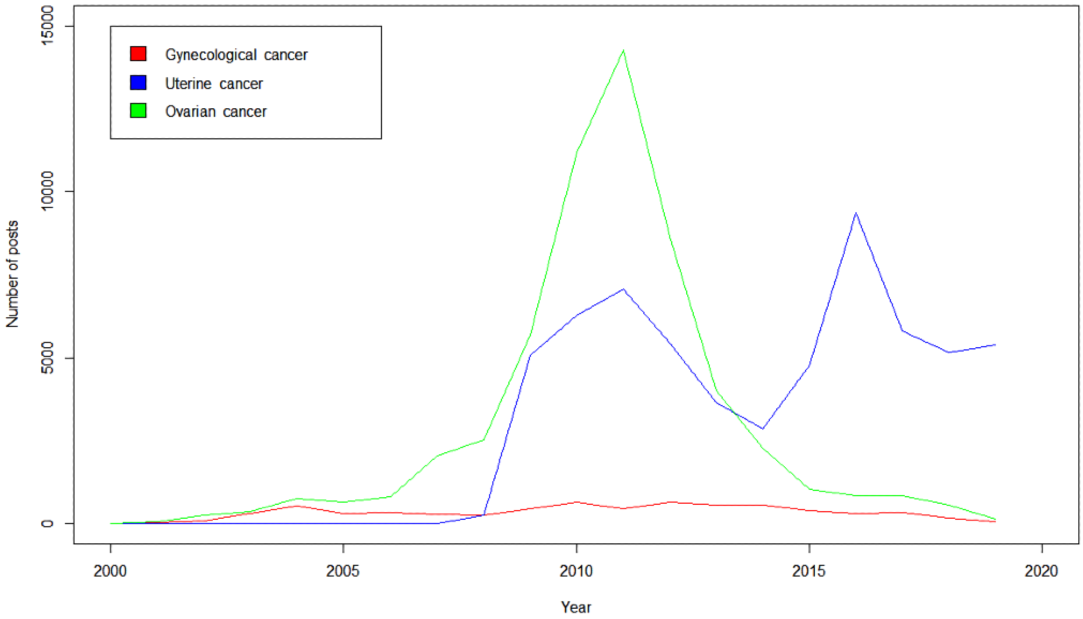


**Figure 2.** Prevalence of psychosocial themes over time.

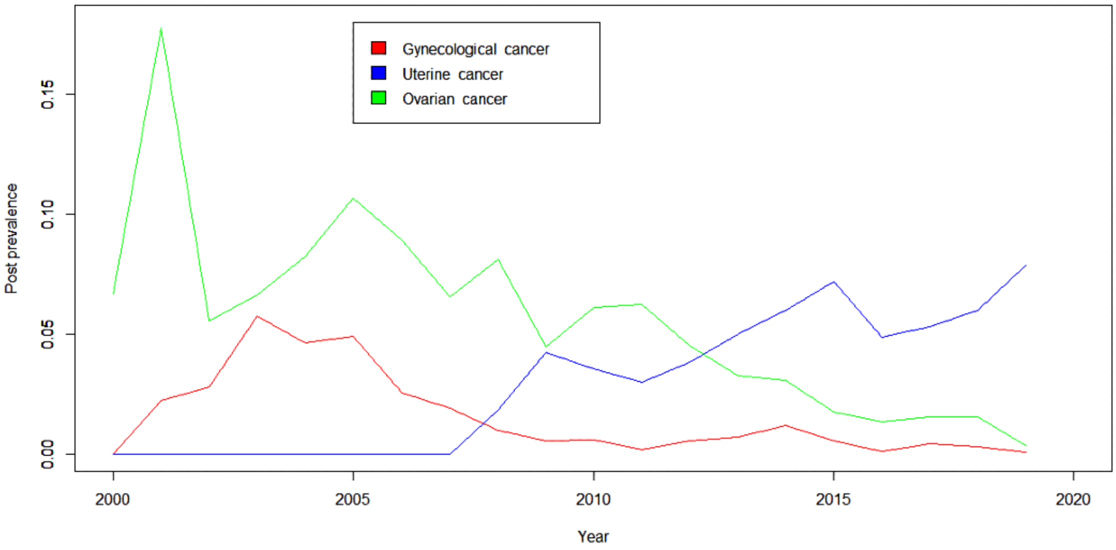
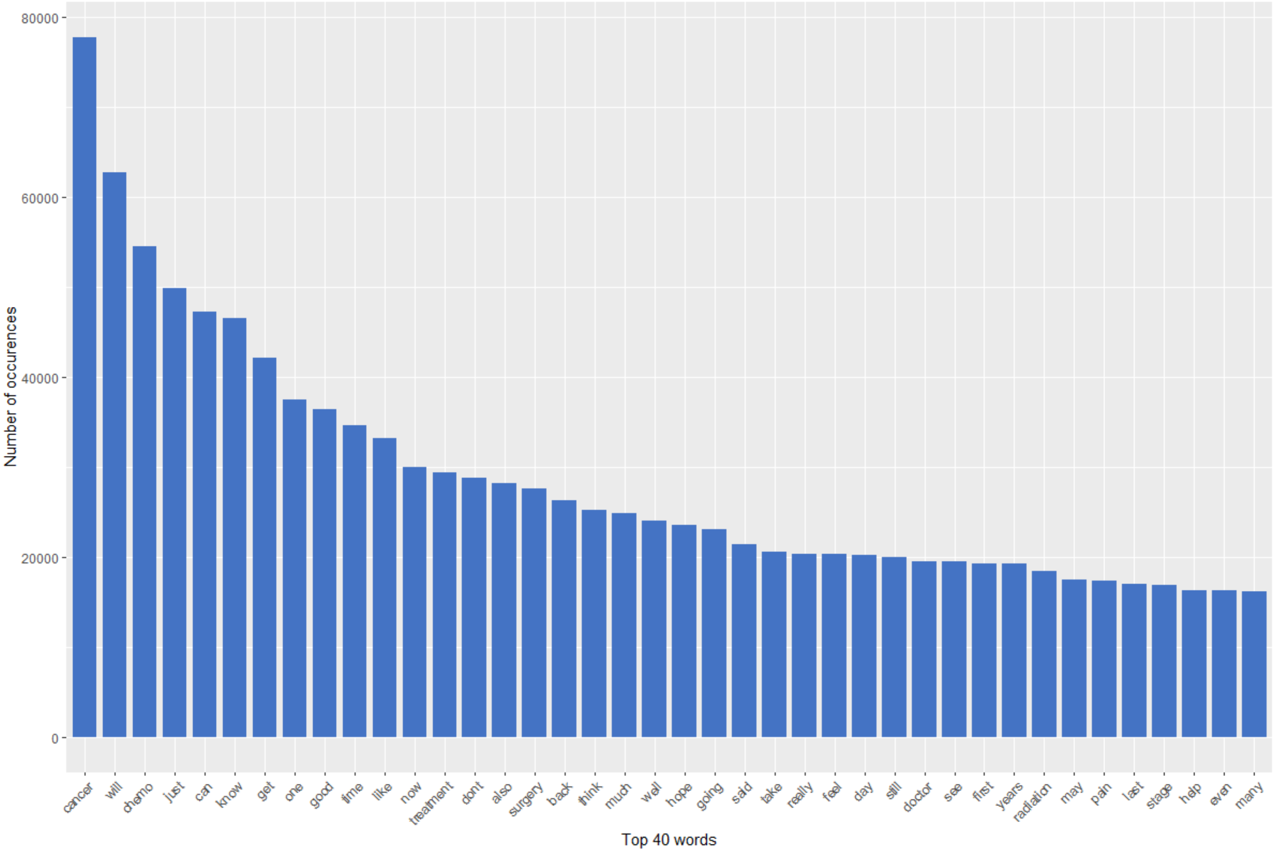


**Figure 3.** Most frequently used words in posts.



## Qualitative Analysis

The 20 posts included in our proof-of-concept qualitative analysis included 238 quotes that were determined to be of interest. These 20 posts came from 8 different users who were either survivors or caregivers of patients with ovarian or uterine cancer. Of the 20 posts, most concerns that were related to the predetermined themes of interest in the *Surveillance and Management of Psychosocial Effects Domain* were related to coping with distress and worry, family or caregiver relationships, and the financial hardships of cancer. Textbox 2 includes representative quotes that discuss these highlighted themes.

In exploring the qualitative aspects of the observed psychosocial posts, additional nuances came to light. A notable amount of the discussion that was related to the psychosocial themes of distress and worry was related to concerns regarding the disease returning, progressing, or metastasizing. This gave us insight into the theme of coping as well, where we observed two salient approaches: taking an active role in care and the role of spirituality. Coping by taking an active role in care was evident as posters requested opinions on therapeutic options, shared and interpreted primary literature with one another, asked for help devising questions for their physician, and brainstormed strategies to self-advocate. A distinct element of this was encouraging others to seek a second opinion, as demonstrated by a uterine cancer discussion board poster:

> *A second opinion may prove lifesaving for you. There are options out there for you in spite of this disappointing news, so don't let a busy doctor write you off. Don't write yourself off either. If you have to go down, go down fighting. Do it for yourself, do it for your husband, and do it for those dreams you have.*

In addition to augmenting self-efficacy, another component of coping was spirituality. Many shared the way that faith provided them reassurance in adjusting to a terminal diagnosis and accepting one's mortality. A patient with metastatic ovarian cancer shared the following:

> *Contrary to what others may choose to believe, while I know that God can and does heal many, He isn't obligated to do so. And I'm not here to challenge anyone else's beliefs, just telling you where I come from as it relates to my own Stage IV diagnosis.*

Survivors also frequently discussed the relationships they had with family and caregivers, with 1 important element being focused on helping loved ones adjust emotionally to the patient's cancer diagnosis, treatment, and prognosis. A survivor of ovarian cancer created a post on how spouses, parents, adult children, friends, and colleagues may support a loved one with cancer, which included (1) rallying the support network to delegate tasks and increase overall support, (2) remembering to take care of oneself so that they are better able to support the patient, and (3) attending support groups for caregivers. Another striking post was from the husband of a woman with ovarian cancer who asked for advice on how to speak to his wife about her cancer and prognosis while she anecdotally was in denial. A survivor replied, suggesting the following:

> *tell her you're researching her cancer. If she asks more, tell her more. If she wants to live in denial,*

> *that's her choice...if she becomes curious enough to ask more, that could be an opening for a deeper conversation. Ideally the two of you need to be able to face the truth together. She will be the one suffering the physical and emotional aspects of this diagnosis. You will certainly be suffering emotionally...and believe me, she wants to live as much as you want her to live...As a caregiver, I know your heart aches.*

Another theme that was prevalent and provided interesting insights into the lived experience of survivorship was the intersection of financial toxicity and insurance status. Conversations were particularly focused on accessing treatment options. A survivor of uterine cancer posted the following:

> *During all this I'm going though, now my insurance is denying paying for the Avastin. I have no idea how much it would cost out of pocket.*

Another survivor replied with the following advice:

> *They will just deny almost as a matter of course unless they're disputed. People are denied all the time and get it overturned after some dispute...They start off with denial and then see if the doctor really means it or not...You really really really have to be on your doctor's office to find out exactly what was submitted, exactly what your insurance requires and exactly why it was denied, and getting your doctor to fight. You shouldn't have to...but you really really really do.*

This provided insights into the experiences that patients may be facing, which affect their quality of life outside of cancer itself.

As demonstrated, the qualitative analysis provided insights into *how* survivors discussed particular themes of interest and gave insights into how these themes are related to one another. Notably, the vast majority of the analyzed quotes from the 20 posts touched on >1 theme in the *Surveillance and Management of Psychosocial Effects Domain*. The least commonly observed psychosocial survivorship themes of interest were related to adherence to treatment, educational and occupational hardships, and social withdrawal.

**Textbox 2.** Quotes from the qualitative analysis.

**Themes and quotes from discussion board posts**

**Coping**

"I'm 42. I weep for the years I'm likley to lose to this cancer and at the rate it's growing I fear I only have months now."

**Distress and worry**

"I speak as a Stage IV cancer patient, and know the effects I have already suffered through, and I'm not about to 'try one more' for only 3 more months of survival, which would not be absent side effects which are many, during that time. I speak only for myself, but my mind is made up!"

**Psychological aspect of physical effects**

"Often it's quite debilitating. And we may even despair at times when side effects seem intolerable."

**Insurance status and financial toxicity**

"Boy what a pain in the butt. I'm sorry you have to fight this money stuff. You have enough to worry about. Boy don't they know how urgent and important this is. At the same time this just shows what a great fighter you are. I really hope you get this cleared so you can get the treatments you need."

**Family and caregiver relationships**

"[as a caregiver], allow them to vent, don't take it personally, know it's the confusion and disease that's talking and all of their fears...Be strong. Don't borrow trouble, but be realistic to the prognosis and day-by-day plan, attitude is huge in the battle and living in the now and positivity is key"

### Topic Modeling

Although LDA was able to separate the posts into different themes, the resulting themes were broad. Investigating a model that discovered 16 rather than 8 themes led to more specific topics, although many were deemed random and irrelevant; therefore, analysis of the 8-theme model was continued. The 8 themes that LDA discovered from greatest theme prevalence to least theme prevalence were support, treatment side effects, diagnosis, research and clinical trials, treatment, ovarian cancer, help, and time. These findings are summarized in Multimedia Appendix 1, as well as the keywords the topic modeling paired together for each theme. Notably, there was an overlap of 2 keywords among the LDA-generated keywords and the keywords developed using our primary mixed methods methodology: *sleep* within the "treatment side effects" discovered theme and *work* within the "time" discovered theme.

## Discussion

### Principal Findings

This study was conducted to apply a novel methodology that was developed to examine a large web-based, narrative-based data set using mixed qualitative and quantitative methods. This methodology offers both the ability to describe and leverage extremely large quantities of data using quantitative techniques while simultaneously guiding and streamlining qualitative analysis. We demonstrated the utility of this methodology in the context of the previously published Quality of Cancer Survivorship Care Framework, with a focus on the framework's *Surveillance and Management of Psychosocial Effects Domain*. We were able to observe both the number of posts made per year across each discussion board, as well as the trends in particular themes of interest across time since the discussion boards' inception. This provided us with a broad overview of engagement on the platform. Quantitative analysis revealed that survivors frequently discussed the role of friends and family in care, as well as fatigue, the impact of interpersonal relationships, and health insurance status. The most frequently used words that survivors wrote in their posts were related to diagnosis and treatment. The most frequently used words related to the themes that we were interested in from the framework's *Surveillance and Management of Psychosocial Effects Domain* were "family," "hope," and "help." Qualitative analysis also demonstrated that survivors frequently discussed coping, worry, distress, and family and caregiver relationships, as well as insurance and financial aspects of care. The qualitative analysis provided deeper insights into *how* survivors were affected by these themes and provided greater insights into the nuances of survivors' unmet concerns.

### Comparison With Prior Work

Social media and discussion-based platforms can provide invaluable information from people affected by cancer and other health conditions; however, the magnitude of data available in social media posts poses a barrier to its use in research [11,28]. Although there are other automated methods to thematically categorize large quantities of written text data in addition to those we have presented in this paper, such as topic modeling [16,17], they often use computer software to generate and cluster themes that the software deems to be of importance. Unfortunately, as we experienced in our analysis, these themes are often not meaningfully relevant enough to answer a specific research question or explore known themes. As we desired to explore the psychosocial aspects of survivorship, the themes that LDA discovered were not specific enough to meaningfully inform our research question. To that end, our methodology ensures that themes of interest are captured while still being an automated process. Another approach to thematically analyzing large quantities of text-based data, which ensures that themes of interest are accurately captured, involves manually extracting data by hand [11-13,15]. However, manual abstraction is often not feasible when wishing to analyze the entirety of the great magnitudes of data that are available on web-based platforms. The methodology we present in this report offers the efficiency of automated thematic analysis while retaining the accuracy and thoughtfulness of manual data abstraction.

The methodology presented in this report may be used to provide clinicians and researchers with invaluable insights, opinions, and suggestions made by survivors of cancer themselves in an efficient and low-cost manner. The candid conversations among survivors publicly available on discussion boards and social media platforms may inform future survivorship efforts and programs, as they reveal the honest and spontaneous concerns, attitudes, and preferences of survivors. This report contributes to the growing body of knowledge extracted from the ACS discussion boards, including a prior publication from our group [15,29]. Additional articles support the richness of the opportunity to explore the perspectives of survivors of cancer through ACS discussion board posts, as demonstrated by Fallon et al [30], who found that 25% of ACS Cancer Survivors Network users return to the site at least monthly, with most interacting with the discussion boards. Harnessing the candid words of survivors will provide survivorship initiatives with an understanding of what is most important to survivors and could perhaps inform clinicians of what is currently missing in survivorship care.

Prior work has explored the unique survivorship needs of those with gynecological cancers. DeRooij et al [31] explored the unmet survivorship needs of patients with gynecological cancers from the perspective of patients, caregivers, and health care providers using semistructured interviews. Almost all participants wished to receive more resources on the side effects of treatment, the anticipated follow-up plan, and psychosocial support. Our analysis contributes to the understanding of what psychosocial supports could be most useful to survivors of gynecological cancer and their caregivers. In fact, Beesley et al [32] distributed a survey to determine the number of survivors of gynecological cancer who had unmet survivorship needs and found that 43% did, with most of their concerns being related to the psychosocial aspects of survivorship. The most common unmet needs included fear of cancer progressing, helping friends and family adjust to their cancer diagnosis, future unknowns, fatigue, and being unable to perform the tasks and activities they enjoyed previously [32]. These findings are very similar to ours, particularly as we found that the greatest worries that survivors discussed on the ACS discussion boards were related to fear of their cancer progressing, metastasizing, or returning. Moreover, the concerns regarding how cancer affects loved ones were salient as multiple posts were related to giving advice on how to help caregivers adjust to their loved one's diagnosis or how caregivers can best support their loved ones with cancer. Together, our work complements and further informs the available literature on the unmet psychosocial needs of survivors of gynecological cancer.

### Limitations

Although our analysis is strengthened by the data set that included 20 years of posts from the ACS discussion boards, we acknowledge that the cancer care and survivorship needs of patients have greatly changed over this time. Thus, the findings must be independently evaluated to determine how they may be applied to current and future survivorship initiatives. As the qualitative analysis served as a proof of concept, we wish to emphasize that the qualitative analysis of this project is not a comprehensive analysis of all of the posts; rather, it demonstrates how this technique can be used to discuss *how* particular topics are discussed. Thus, the quotes explored in the qualitative analysis may not be reflective of all posts in the data set. Moreover, we acknowledge the limited diversity in the authors of the 20 posts qualitatively analyzed.

### Conclusions

Internet-based discussions among survivors of gynecological cancer offer valuable insights into the unmet psychosocial survivorship needs that can be addressed in future survivorship initiatives. Most often, survivors of gynecological cancer discussed the role of friends and family in care, as well as fatigue, the effect of cancer on interpersonal relationships, and health insurance status, as discovered through the quantitative phase of the analysis. The complementary qualitative analysis informed how these themes affect survivors, showing the specific gaps in survivorship care that may be addressed. This informative and customizable methodology may continue to be applied across clinical settings and patient populations, harnessing the abundance of patient-generated and patient-centered internet data for empirical inquiry.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Themes and keywords discovered via latent Dirichlet allocation.
[PNG File , 75 KB - jmir_v24i9e37757_app1.png ]

## References

1. American Cancer Society. American Cancer Society Cancer Statistics 2021 Report. J Nucl Med 2021 Mar;62(3):12N [FREE Full text] [Medline: 33622967]
2. National Research Council, Institute of Medicine, National Cancer Policy Board, Committee on Cancer Survivorship: Improving Care and Quality of Life. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC, USA: National Academies Press (US); 2006.
3. NCI Dictionaries - Survivorship. National Cancer Institute. 2022. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/survivor [accessed 2022-09-08]
4. NCCN Survivorship Guidelines: Survivorship Care for Cancer-Related Late and Long-Term Effects. National Comprehensive Cancer Network. 2020. URL: https://www.nccn.org/patients/guidelines/content/PDF/survivorship-crl-patient.pdf [accessed 2022-09-08]
5. Ontario C. Follow-Up Model of Care for Cancer Survivors. Guidelines 2019 [FREE Full text]
6. American Cancer Society Guidelines for Cancer Survivors. American Cancer Society. 2022. URL: https://www.cancer.org/health-care-professionals/american-cancer-society-survivorship-guidelines.html [accessed 2022-09-08]
7. Guidelines on Survivorship Care. American Society of Clinical Oncology. 2022. URL: https://www.asco.org/news-initiatives/current-initiatives/cancer-care-initiatives/survivorship-compendium/guidelines [accessed 2022-09-08]
8. Nekhlyudov L, Mollica MA, Jacobsen PB, Mayer DK, Shulman LN, Geiger AM. Developing a quality of cancer survivorship care framework: implications for clinical care, research, and policy. J Natl Cancer Inst 2019 Nov 01;111(11):1120-1130 [FREE Full text] [doi: 10.1093/jnci/djz089] [Medline: 31095326]
9. NCI Dictionaries - Psychosocial Effects. National Cancer Institute. 2022. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/psychosocial [accessed 2022-09-08]
10. Falisi AL, Wiseman KP, Gaysynsky A, Scheideler JK, Ramin DA, Chou WY. Social media for breast cancer survivors: a literature review. J Cancer Surviv 2017 Dec;11(6):808-821. [doi: 10.1007/s11764-017-0620-5] [Medline: 28601981]
11. Gao RW, Smith JD, Malloy KM. Head and neck cancer and social media: the patient experience and cancer survivorship. Laryngoscope 2021 Apr;131(4):E1214-E1219. [doi: 10.1002/lary.29074] [Medline: 32886368]
12. Sutton J, Vos SC, Olson MK, Woods C, Cohen E, Gibson CB, et al. Lung cancer messages on Twitter: content analysis and evaluation. J Am Coll Radiol 2018 Jan;15(1 Pt B):210-217. [doi: 10.1016/j.jacr.2017.09.043] [Medline: 29154103]
13. Mikal JP, Beckstrand MJ, Parks E, Oyenuga M, Odebunmi T, Okedele O, et al. Online social support among breast cancer patients: longitudinal changes to Facebook use following breast cancer diagnosis and transition off therapy. J Cancer Surviv 2020 Jun;14(3):322-330. [doi: 10.1007/s11764-019-00847-w] [Medline: 31897878]
14. Erfani SS, Abedin B, Daneshgar F. A qualitative evaluation of communication in Ovarian Cancer Facebook communities. In: Proceedings of the International Conference on Information Society. 2013 Presented at: i-Society '13; June 24-26, 2013; Toronto, Canada p. 270-227.
15. Smith A, Fogarasi M, Lustberg MB, Nekhlyudov L. Perspectives of adolescent and young adult cancer survivors: review of community-based discussion boards. J Cancer Surviv (forthcoming) 2021 Sep 02. [doi: 10.1007/s11764-021-01098-4] [Medline: 34471949]
16. Bisgin H, Liu Z, Fang H, Xu X, Tong W. Mining FDA drug labels using an unsupervised learning technique--topic modeling. BMC Bioinformatics 2011 Oct 18;12 Suppl 10:S11 [FREE Full text] [doi: 10.1186/1471-2105-12-S10-S11] [Medline: 22166012]
17. Jones J, Pradhan M, Hosseini M, Kulanthaivel A, Hosseini M. Novel approach to cluster patient-generated data into actionable topics: case study of a Web-based breast cancer forum. JMIR Med Inform 2018 Nov 29;6(4):e45 [FREE Full text] [doi: 10.2196/medinform.9162] [Medline: 30497991]
18. Lee YJ, Park A, Roberge M, Donovan H. What can social media tell us about patient symptoms: a text-mining approach to online ovarian cancer forum. Cancer Nurs 2022;45(1):E27-E35. [doi: 10.1097/NCC.0000000000000860] [Medline: 32649337]

19. Vedel I, Kaur N, Hong Q, El Sherif R, Khanassov V, Godard-Sebillotte C, et al. Why and how to use mixed methods in primary health care research. Fam Pract 2019 May 23;36(3):365-368 [FREE Full text] [doi: 10.1093/fampra/cmy127] [Medline: 31120129]
20. Creswell JW. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 4th edition. Thousand Oaks, CA, USA: Sage Publications; 2013.
21. Pluye P, Gagnon MP, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. Int J Nurs Stud 2009 Apr;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
22. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022.
23. Rōder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 Presented at: WSDM '15; February 2-6, 2015; Shanghai, China p. 399-408. [doi: 10.1145/2684822.2685324]
24. Sievert C, Shirley K. LDAvis: a method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces. 2014 Presented at: WS '14; June 27, 2014; Baltimore, MD, USA p. 63-70. [doi: 10.3115/v1/w14-3110]
25. Office for Human Research Protections (OHRP). 2022. URL: https://www.hhs.gov/ohrp/regulations-and-policy/index.html [accessed 2022-09-08]
26. Boddy CR. Sample size for qualitative research. Qual Mark Res 2016 Sep 12;19(4):426-432. [doi: 10.1108/qmr-06-2016-0053]
27. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. Thousand Oaks, CA, USA: Sage Publications; 2013.
28. Berry MW, Mohamed A, Yap BW. Supervised and Unsupervised Learning for Data Science. Cham, Switzerland: Springer; 2020.
29. Portier K, Greer GE, Rokach L, Ofek N, Wang Y, Biyani P, et al. Understanding topics and sentiment in an online cancer survivor community. J Natl Cancer Inst Monogr 2013 Dec;2013(47):195-198. [doi: 10.1093/jncimonographs/lgt025] [Medline: 24395991]
30. Fallon EA, Driscoll D, Smith TS, Richardson K, Portier K. Description, characterization, and evaluation of an online social networking community: the American Cancer Society's Cancer Survivors Network®. J Cancer Surviv 2018 Oct;12(5):691-701. [doi: 10.1007/s11764-018-0706-8] [Medline: 30083972]
31. de Rooij BH, Thomas TH, Post KE, Flanagan J, Ezendam NP, Peppercorn J, et al. Survivorship care planning in gynecologic oncology-perspectives from patients, caregivers, and health care providers. J Cancer Surviv 2018 Dec;12(6):762-774 [FREE Full text] [doi: 10.1007/s11764-018-0713-9] [Medline: 30209681]
32. Beesley V, Eakin E, Steginga S, Aitken J, Dunn J, Battistutta D. Unmet needs of gynaecological cancer survivors: implications for developing community support services. Psychooncology 2008 Apr;17(4):392-400. [doi: 10.1002/pon.1249] [Medline: 17680554]

## Abbreviations

**ACS:** American Cancer Society
**CV:** metric of coherence
**LDA:** latent Dirichlet allocation

*Edited by A Mavragani; submitted 05.03.22; peer-reviewed by Y Xie, S El kefi, W Ceron, R El Sherif; comments to author 27.06.22; revised version received 17.07.22; accepted 11.08.22; published 20.09.22.*

*Please cite as:*
*Adams EJ, Tallman D, Haynam ML, Nekhlyudov L, Lustberg MB*
*Psychosocial Needs of Gynecological Cancer Survivors: Mixed Methods Study*
*J Med Internet Res 2022;24(9):e37757*
*URL: https://www.jmir.org/2022/9/e37757*
*doi:10.2196/37757*
*PMID:36125848*

Original Paper

# Newborn Skin Maturity Medical Device Validation for Gestational Age Prediction: Clinical Trial

Zilma Silveira Nogueira Reis[1,2,3*], MD, PhD; Roberta Maia de Castro Romanelli[4], MD, PhD; Rodney Nascimento Guimarães[2*], PhD; Juliano de Souza Gaspar[2], CMP, PhD; Gabriela Silveira Neves[5*], PT, MSc; Marynea Silva do Vale[6*], MSc, MD; Paulo de Jesus Nader[7*], MD, PhD; Martha David Rocha de Moura[8*], MSc, MD; Gabriela Luíza Nogueira Vitral[4*], MSc; Marconi Augusto Aguiar dos Reis[4*], MSc, MD; Marcia Margarida Mendonça Pereira[6*], MD; Patrícia Franco Marques[6*], MD; Silvana Salgado Nader[7*], MSc, MD; Augusta Luize Harff[7*], MD; Ludmylla de Oliveira Beleza[8*], PT; Maria Eduarda Canellas de Castro[8*], MD; Rayner Guilherme Souza[2*], MS; Gisele Lobo Pappa[3,9*], CMP, PhD; Regina Amélia Pessoa Lopes de Aguiar[10*], MD, PhD

[1]Health Informatics Center, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[2]Department of Gynecology and Obstetrics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[3]Center for Artificial Intelligence, Innovation and Health, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[4]Child and Adolescent Health, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[5]Hospital Sofia Feldman, Belo Horizonte, Brazil
[6]Maternal and Child Unit, University Hospital, Universidade Federal do Maranhão, São Luis, Brazil
[7]University Hospital of Canoas, Universidade Luterana do Brasil, Canoas, Brazil
[8]Hospital Materno Infantil de Brasília, Distrito Federal, Brazil
[9]Computer Science Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[10]State Health Department of Minas Gerais, Belo Horizonte, Brazil
*these authors contributed equally

**Corresponding Author:**
Zilma Silveira Nogueira Reis, MD, PhD
Health Informatics Center
Universidade Federal de Minas Gerais
Avenue Professor Alfredo Balena 190
Belo Horizonte, 30130100
Brazil
Phone: 55 3134099648
Email: zilma.medicina@gmail.com

## Abstract

**Background:** Early access to antenatal care and high-cost technologies for pregnancy dating challenge early neonatal risk assessment at birth in resource-constrained settings. To overcome the absence or inaccuracy of postnatal gestational age (GA), we developed a new medical device to assess GA based on the photobiological properties of newborns' skin and predictive models.

**Objective:** This study aims to validate a device that uses the photobiological model of skin maturity adjusted to the clinical data to detect GA and establish its accuracy in discriminating preterm newborns.

**Methods:** A multicenter, single-blinded, and single-arm intention-to-diagnosis clinical trial evaluated the accuracy of a novel device for the detection of GA and preterm newborns. The first-trimester ultrasound, a second comparator ultrasound, and data regarding the last menstrual period (LMP) from antenatal reports were used as references for GA at birth. The new test for validation was performed using a portable multiband reflectance photometer device that assessed the skin maturity of newborns and used machine learning models to predict GA, adjusted for birth weight and antenatal corticosteroid therapy exposure.

**Results:** The study group comprised 702 pregnant women who gave birth to 781 newborns, of which 366 (46.9%) were preterm newborns. As the primary outcome, the GA as predicted by the new test was in line with the reference GA that was calculated by using the intraclass correlation coefficient (0.969, 95% CI 0.964-0.973). The paired difference between predicted and reference GAs was −1.34 days, with Bland-Altman limits of −21.2 to 18.4 days. As a secondary outcome, the new test achieved 66.6%

(95% CI 62.9%-70.1%) agreement with the reference GA within an error of 1 week. This agreement was similar to that of comparator-LMP-GAs (64.1%, 95% CI 60.7%-67.5%). The discrimination between preterm and term newborns via the device had a similar area under the receiver operating characteristic curve (0.970, 95% CI 0.959-0.981) compared with that for comparator-LMP-GAs (0.957, 95% CI 0.941-0.974). In newborns with absent or unreliable LMPs (n=451), the intent-to-discriminate analysis showed correct preterm versus term classifications with the new test, which achieved an accuracy of 89.6% (95% CI 86.4%-92.2%), while the accuracy for comparator-LMP-GA was 69.6% (95% CI 65.3%-73.7%).

**Conclusions:** The assessment of newborn's skin maturity (adjusted by learning models) promises accurate pregnancy dating at birth, even without the antenatal ultrasound reference. Thus, the novel device could add value to the set of clinical parameters that direct the delivery of neonatal care in birth scenarios where GA is unknown or unreliable.

**International Registered Report Identifier (IRRID):** RR2-10.1136/bmjopen-2018-027442

***(J Med Internet Res 2022;24(9):e38727)*** doi:10.2196/38727

**KEYWORDS**

gestational age; prematurity; childbirth; skin physiological phenomena; machine learning; equipment and supplies; pregnancy; reproductive health; pregnant; skin; age; medical; device; newborn; baby; trimester; therapy; learning model; ultrasound

# *Introduction*

## Background

Being born before 37 weeks of gestation, which is preterm birth, is the leading cause of childhood mortality. The global preterm birth rate is approximately 11%, with a particularly high frequency in low- and middle-income countries, in association with maternal education, race, and ethnic origin [1]. However, adverse neonatal outcomes affect newborns unevenly according to the birth scenario and gestational age (GA) [1]. Mortality on the first day of life is 30 times higher in low- and medium-income countries than in high-income countries [2]. The first step in caring for preterm newborns is to identify them, which remains challenging in scenarios with scarce resources [3]. An accurate assessment of preterm newborns at birth can allow practical decisions regarding support, such as keeping the lungs airing, keeping the body warm, regulating metabolism and nutrition, or making decisions to transfer them to an intensive care unit, otherwise avoiding unnecessary interventions for term newborns [4]. Preterm neonates are more prone to death or survival with neurological sequelae. Long-term surviving preterm infants are at risk of death before the age of 5 years and at risk of presenting cognitive and motor sequelae compared with term infants [1]. The need to pinpoint early risks at birth faces the issues of reduced early access to antenatal care and a lack of access to high-cost technologies for pregnancy dating, such as obstetric echography in resource-constrained settings [5].

Some pregnancy-dating troubles arise from antenatal care. Government policies and best practices advise pregnant women to plan pregnancy to include early access to antenatal care for pregnancies to be safely monitored until birth [6]. However, many barriers to covering all pregnancies and births with due care have not been overcome, particularly in scenarios lacking well-equipped facilities [7]. Early obstetric ultrasound currently offers the best method for the establishment of GA [8]. However, lack of access to high-cost equipment, poor training, lack of skills of health professionals, and delayed antenatal care limit pregnancy dating and, consequently, detection of prematurity [5,9]. In addition, GA calculation based on the last menstrual period (LMP) is affected by memory bias, hormone-based contraception, and breastfeeding [10]. After-birth approaches for pregnancy dating, which are also extensively used, rely on professional skills for physical and neurological maturity assessment. Nevertheless, maturity scores have failed in terms of reproducibility and accuracy [3]. Meanwhile, birth weight is a helpful predictor of risk to the newborn and not GA, as size at birth results from the dynamic process of past intrauterine growth beyond the gestation length [9,11].

Reliable pregnancy dating has an impact on measuring the global burden of preterm birth and the associated risks [3,12]. Improving preterm birth outcomes requires accurate assessment of GA to instruct timely decision-making regarding neonatal care [10]. Approaches for the enhancement of the accuracy of pregnancy dating through more accurate and accessible technologies can improve pregnancy outcomes and neonatal survival rates [8,13]. Health technology development is critical for supporting health care systems. Medical devices and digital health technologies have brought innovative solutions with the potential to save lives [14], mitigating quality gaps among disparate health care scenarios [15]. Furthermore, digital health technologies have the potential to impact the equality of health care, creating new landscapes of opportunities, such as application of data science to improve prediction models [16]. Currently, computer science has advanced, with improvements to medical practice, detecting patterns by processing data sets through layered mathematical models [17], and fostering the skills and competences of professionals in support of the best health care decisions [14].

The new test explored in this study is an innovative approach used to estimate GA based on the photobiological properties of the newborn's skin and by learning predictive models enhanced with clinical variables [18]. It being usable as a medical device, we developed this technology to easily assist health professionals in the care of newborns whenever the pregnancy dating is unknown or doubtful, adding relevant information for classification and better management of the newborn.

## Objective

This study aimed to validate a new medical device used to assess GA through the photobiological model of skin maturity adjusted to clinical data and to determine its accuracy in detecting preterm newborns. We tested the hypothesis of equivalence between GA measured by this new test and by pregnancy-dating comparators calculated using ultrasound examinations and the LMP.

## Methods

### Study Design and Participants

This study was a multicenter, prospective, intention-to-diagnosis clinical trial investigation with a single group, single-blinded, and single-arm, using a reference standard. This paper adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prediction or Diagnosis for completeness and clarity [19]. Intention-to-diagnosis is a method for prospective studies in which all participants are considered in the statistical analysis, allowing us to reach unbiased conclusions regarding the effectiveness of an intervention [20]. To assess the risk of bias and applicability, the development and validation methods followed guidance from the Prediction Model Risk of Bias Assessment Tool [21]. The clinical trial protocol was disclosed in the World Health Organization's International Clinical Trial Platform—Brazilian Clinical Trials (registered under trial number RBR-3f5bm5).

This report examined the primary and secondary outcomes of data concerning GA prediction and clinical safety of the novel device. Secondary outcomes related to lung maturity prediction are currently under analysis for further publication. The following five Brazilian urban referral centers for high-complexity perinatal care took part in the study: Clinical Hospital—Universidade Federal de Minas Gerais (as coordinator), Minas Gerais State; Sofia Feldman Hospital—Minas Gerais State; Hospital da Universidade Luterana do Brasil—Rio Grande do Sul State; Hospital Materno-infantil de Brasília—Federal District; and Hospital Universitário da Universidade Federal do Maranhão—Maranhão State.

A prospective concurrent and sequential process enrolled newborns during the first 24 hours of life. The first enrollment occurred on January 2, 2019, and the last occurred on May 30, 2021. Eligibility criteria, participants' timeline, and procedures followed the research clinical protocol [22]. In short, we assessed the skin maturity of live newborns with at least ≥24 weeks of GA. All had reports of antenatal ultrasound, one from 7 to 13 weeks and 6 days and the other from 14 to 23 weeks and 6 days of gestation. Anhydramnios, hydrops, congenital skin diseases, or chorioamnionitis were the exclusion criteria, owing to their potential to modify the skin structure.

### Procedures

The coordinating unit trained 15 health professional examiners following good clinical practice as set forth by the Brazilian Regulatory Health Agency's recommendations. Standard operating procedures were mandatory to guide the enrollment process, skin assessment, and data collection [22]. Clinical information was collected through structured questionnaires, using a software program dedicated to this project. The framework of the clinical variables and skin acquisitions is available in Multimedia Appendix 1. Textual information was saved on a tablet with internet access, individually associated with the respective skin assessment acquired using the medical device [23].

An automated algorithm in the data collection system [24] blinded to the examiner calculated the reference GA. Established rules for redating GA at birth provided our reference for GA using data from the ultrasound reports or antenatal care books or other clinical document [8]. For data curation, the investigator's data entries were confronted with information from photographed digital images of clinical documents. In the case of multiple birth gestations with different ultrasonographic crown-rump length values, the average of each embryo or fetal value was considered. A double-check system, paper-based and electronic, allowed verification of the reliability and validity of clinical data as well as skin reflectance acquisition. In addition, the data quality of antenatal pregnancy dating was evaluated by comparing the frequency of days in dates of LMP, as they should be random with no preference for digits. For this purpose, in cases of multiple gestations, we retained only the first twin information for the day digit evaluation.

### Intervention

The intervention in this clinical trial was a test performed with a novel device that processes the backscattered signal acquired from the skin of the newborn's sole with clinical variables to predict the GA. Its development includes steps from the workbench to clinical experimentation, as described earlier [18]. Similarly, we previously analyzed the best body position to assess skin reflectance for pregnancy dating and environmental influences such as humidity, temperature, ambient light, and the newborn's skin hue [18,25]. Regarding the characteristics of the components, wavelengths from 400 nm to 1200 nm of the light emitter placed the safety level of this medical device in class II (noninvasive and medium risk), according to the regulatory agency in Brazil. When the light-emitting sensor touches the skin of the sole for a few seconds, it triggers 10 automated measurements. The device-emitted error warning signals were caused by the involuntary movement of the newborn or examiner under the input of ambient light by the sensor; these were events that required a new attempt [18]. The device output was blinded to the examiners. The reliability of skin reflection acquisition was assessed during the certification visit of a senior researcher in the collaborating units (Multimedia Appendix 2 [26]).

Skin assessments occurred with the newborn inside incubators, incubators-radiant warmers, warming pad-bassinets, standard cribs, or in the mother's lap to ensure minimum manipulation and to avoid unbalancing the clinical conditions. The sensor touched the sole 3 times, following complete disinfection with alcohol. A total of 14 minimum viable products were produced in this study (Figure 1). At the beginning and end of the clinical trial, the irradiance emitted by each device and the reflection against a standard white wavelength calibration standard provided values for calibration. The adjusted value was the raw value of the acquisition divided by the irradiance of the light-emitting diode of each device.

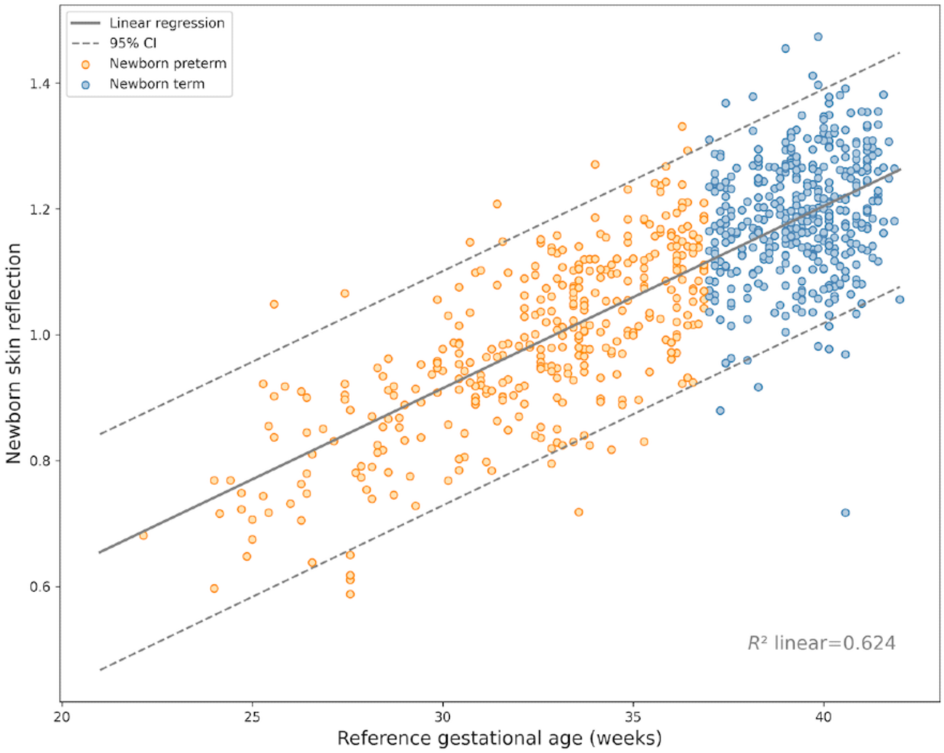
The device used an algorithm to predict GA, as previously described, and was duly patented [18]. We assessed the Pearson coefficient to confirm the correlation between skin reflectance and the reference GA. Skin reflectance had a strong positive correlation with the reference GA ($r$=0.79, $P$<.001; Figure 2).

The standalone newborn skin reflectance value was adjusted for clinical variables. The current data set the groundwork for improvements in the model for prediction of GA with the use of machine learning models as part of the research protocol. The analytical pipeline is detailed in Multimedia Appendix 3 [27-29]. The nonlinear machine learning method, Extreme Gradient Boosting (XGBoost) [27], created no more than 50 trees with a maximum depth of 3. The models were validated using a 10-fold cross-validation approach repeated 30 times. Clinical variables used as predictors of the models were available at the time of testing, which is a part of the routine of care. Therefore, they can be used in real scenarios from user input into the medical device interface.

Models' performance with different covariates, including intermediate analysis considering factors such as incubator stay, sex, and jaundice, is presented in Multimedia Appendix 4 [18]. These new tests were performed to validate the elimination of intervenient variables after technological improvements and were added to the current version of the device [18]. Skin acquisition, duly adjusted for antenatal corticosteroid therapy for fetal maturation (ACTFM) exposure information, achieved a coefficient of determination, $R^2$ of 0.732 and a mean absolute error (MAE) of 1.688 weeks (11.8 days). In addition, considering birth weight, the model achieved an even better performance in terms of $R^2$ of 0.878, with an MAE of 1.147 weeks (8.0 days). This new model, with 3 predictive variables, was the one validated in this study. However, 3 GA predictions had ACTFM data imputation by the machine learning model owing to missing information because of failures in the antenatal record available at maternity admission.

**Figure 1.** The new device and its simulated application on a newborn doll.

**Figure 2.** Correlation plot between the skin reflectance of the newborn and the reference gestational age at birth.



### Outcomes

The primary outcome was the agreement between the GA predicted by the device and reference GA. The secondary end point was the accuracy of the device in the identification of preterm newborns, considering thresholds at 37, 32, and 28 weeks of pregnancy. Moreover, the proportion of preterm newborns correctly detected at birth within a 1-week error margin. Another secondary end point was the comparison of differences between predicted GA and GA calculated by a second ultrasound examination after 13 weeks and 6 days of gestation and before 22 weeks via comparator-ultrasound-GA and with the comparator-LMP-GA. This outcome was intended to simulate the performance of the device in scenarios without the reference and to compare the agreement between the established methods for GA calculation and the new test. The safety of the device is still a derived end point which refers to the reporting of unexpected medical events, unintended illness or injury, or adverse clinical signs in newborns, users, or others, regardless of whether they are related to the investigated product. The users answered 9 questions regarding issues with the medical device, after each skin acquisition (Multimedia Appendix 1).

### Statistical Analysis

Descriptive analyses of the newborn's clinical characteristics and intervention measurements were performed. Regarding the primary end point, the agreement among different methods for GA at birth determination was calculated using the intraclass coefficient (ICC) correlation, Bland-Altman intervals, and the paired day difference to reference GA. Regarding the accuracy of the predicted GA by the device in identifying premature newborns, the area under the receiver operating characteristic curve (AUROC) at a CI of 95% described the new test's discrimination and diagnostic parameters. The chi-square test, Mann-Whitney *U* test, and mean paired differences were used to compare interest groups of preterm and term newborns. *P* values of <.05 were considered suggestive of statistical significance. SPSS software (version 19.0; IBM Corporation) was used for statistical analysis of the data.

### Ethics Approval

The local independent ethics review board approved the research protocol, registered under the number CAAE 81347817.6.1001.5149 at the Brazilian National Research Council. In addition, parents signed an informed consent form on behalf of the newborns before participating.

## Results

### Study Design and Participants

Of the 791 potentially eligible newborns, 2 were under Rh alloimmunization during pregnancy, which was considered an exclusion criterion (Figure 3). In this figure, the test is the prediction of GA with the device using the XGBoost algorithm, which includes skin reflectance, birth weight, and ACTFM exposure predictors. The positive sign ("+") represents preterm and negative sign ("−") represents term. Among the 789 newborns who had their skin assessed with the optical probe of the device, 8 had no reference standard to assess the dependent variable, 4 had no antenatal first-trimester ultrasound, 3 had no comparator ultrasound, and 1 had an unsolved digit date error. All 781 newborns who met the eligibility criteria for the clinical trial were included in the analysis.

The study group comprised 702 pregnant women who gave birth to 781 newborns. Despite early access to antenatal care with a median value of 12 (IQR 4) weeks (Table 1), only 296 (42.2%) women met the criteria for reliable LMP among 613 who were able to provide such a date. According to the reference GA at birth, 53.1% (415/781) of newborns were born at term. Among 366 (46.9%) preterm newborns, 235 (30.1%) had a GA at birth of 32 to 37 weeks, 131 (16.8%) had a GA of 28 to 32 weeks, and 42 (5.4%) had a GA of less than 28 weeks. Some newborns (273/781, 35.1%) received ACTFM following local protocols, and in 3 (0.4%), the data were missing. The frequency of abnormal fetal growth classification at birth was 115 (14.7%) in the small for GA group and 59 (7.6%) in the large for GA group. Approximately one-third (280/781, 35.9%) of the newborns were in the intensive care unit at the time of skin assessment.

**Figure 3.** Flow diagram of participants throughout the study with results for the predictive model. GA: gestational age; US: ultrasound.

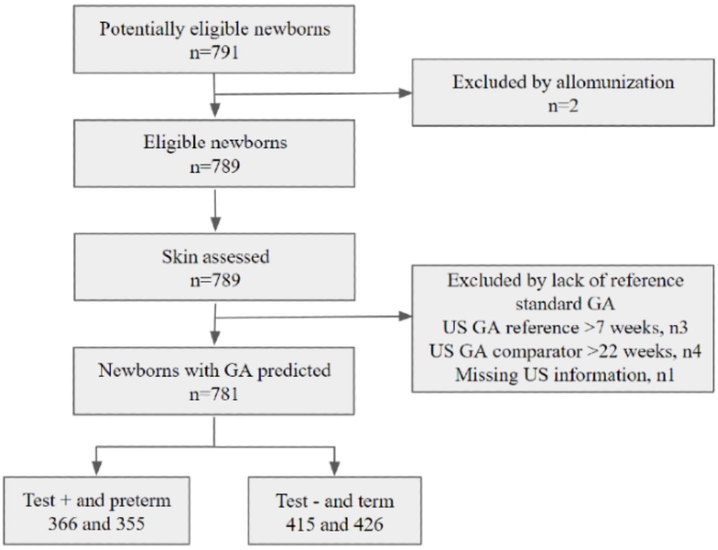


**Table 1.** Baseline characteristics of the pregnancies and newborns.

| Characteristics | Values, n | Statistics |
|---|---|---|
| **Maternal data** | 702 | N/A[a] |
| Maternal age (years), median (IQR) | 702 | 27 (9) |
| First antenatal care assessment (weeks), median (IQR) | 616 | 12 (4) |
| Absent recall of last menstrual period, n (%) | 702 | 89 (12.7) |
| Reliable last menstrual period, n (%) | 613 | 296 (42.2) |
| Diabetes, n (%) | 701 | 103 (14.7) |
| Hypertensive disturbance during pregnancy, n (%) | 702 | 1103 (14.7) |
| ACTFM[b], n (%) | 698 | 273 (35.1) |
| Multiple gestation, n (%) | 702 | 74 (10.5) |
| **Neonatal data** | 781 | N/A |
| Reference gestational age at birth (weeks), median (IQR) | 781 | 37.3 (6.3) |
| Gestational age at the first ultrasound assessment (weeks), median (IQR) | 781 | 10.1 (3.6) |
| Gestational age at the second ultrasound assessment (weeks), median (IQR) | 781 | 19.4 (4.3) |
| ACTFM exposure, n (%) | 777 | 273 (35.1) |
| Major malformation, n (%) | 781 | 8 (1.1) |
| 1-min Apgar score, median (IQR) | 775 | 8 (1) |
| 5-min Apgar score, median (IQR) | 777 | 9 (1) |
| Birth weight (g), median (IQR) | 781 | 2740 (1498) |
| Sex (male), n (%) | 781 | 390 (49.9) |
| Incubator accommodation at skin assessment, n (%) | 781 | 239 (30.6) |
| NICU[c] at skin assessment, n (%) | 781 | 280 (35.9) |
| Jaundice at skin assessment, n (%) | 779 | 255 (32.7) |
| Phototherapy at skin assessment, n (%) | 774 | 32 (4.1) |
| Newborn mortality within first 72 hours, n (%) | 781 | 14 (1.8) |
| Respiratory distress syndrome until 72 hours, n (%) | 781 | 215 (27.5) |
| **Classifications of newborns based on reference gestational age** | | |
| Preterm[d], n (%) | 781 | 366 (46.9) |
| Moderate to late preterm[e], n (%) | 781 | 235 (30.2) |
| Very preterm[f], n (%) | 781 | 89 (11.4) |
| Extremely preterm[g], n (%) | 781 | 42 (5.4) |
| Small for gestational age, n (%) | 781 | 115 (14.7) |
| Appropriate for gestational age, n (%) | 781 | 607 (77.7) |
| Large for gestational age, n (%) | 781 | 59 (7.6) |

[a] N/A: not applicable.

[b] ACTFM: antenatal corticosteroid therapy for fetal maturation.

[c] NICU: neonatal intensive care unit.

[d] Less than 37 weeks.

[e] More than 32 to less than 37 weeks.

[f] More than 28 to less than 32 weeks.

[g] Less than 28 weeks.

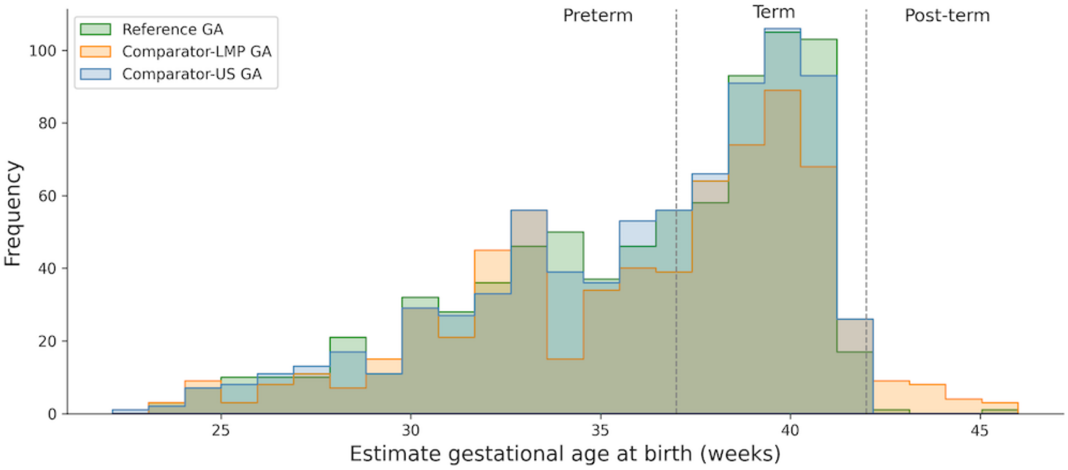
## Procedures: GA at Birth by Established Methods

The distribution of GA as calculated according to the different references corroborated some differences among the established methods of antenatal dating, as shown in the overlapped histogram, in weeks of gestation (Figure 4). In this figure, the red dotted line corresponds to the limit between preterm and term newborns. The green dotted line corresponds to the limit between term and postterm newborns. Reference GA had a median of 37.3 (IQR 6.3) weeks, above that of the comparator-ultrasound-GA, which had a median of 37.1 (IQR 6.1) weeks, *P*<.001 (paired Wilcoxon test). However, when available, the comparator-LMP-GA had a median of 37.4 (IQR 6.8) weeks, similar to the reference GA, *P*=.282 (paired Wilcoxon test). The frequency of preterm birth was 46.9% (366/781), 47.1% (368/781), and 45.6% (310/680) according to the reference GA, comparator-ultrasound-GA, and comparator-LMP-GA, respectively. The frequency of postterm birth was 0.1% (1/781), 0.3% (2/781), and 4% (27/680) with reference to GA, comparator-ultrasound-GA, and comparator-LMP-GA, respectively. On the other hand, the data quality of the LMP recall revealed that the most frequent digit preferences were for days 5 (8.3%), 15 (6.7%), 20 (7.2%), and 25 (4.7%). These frequencies had significant differences when compared with the day adjusted to the reference GA (*P*<.008; Cochran Q test for k-related samples).

Digit preference analysis searched for the tendency of round-numbered days of the menstrual period, considering digits, typically multiples of 5 and 10. This was determined by comparing the observed and expected counts for each day of a month. The Cochran Q test for k-related samples compared the LMP with the day adjusted to the reference GA. We removed duplicate data from twins, and observations on day 31 were removed during the statistical test. The dotted line corresponds to the frequency expected for each day for 30 days per month.

Analyzing the day digit of the LMP informed by the woman, the most frequent digit preferences were for days 5 (8.3%), 15 (6.7%), 20 (7.2%), and 25 (4.7%). These frequencies had significant differences when compared to the day adjusted to the reference GA (*P*=.008). Analyzing the day digit of the LMP adjusted by the second ultrasound examination performed on data after 13 weeks and 6 days of gestation and before 22 weeks (comparator ultrasound), there were no significant differences when compared with the day adjusted to the reference GA (*P*=.20).

**Figure 4.** The distribution of estimated gestational age at birth by the established methods evaluated in this study. GA: gestational age; LMP: last menstrual period; US: ultrasound.



## Primary Outcome: GA Estimation at Birth

The agreement between the predicted GA, reference, and comparators was high considering the ICC (Table 2). Considering the CI of 95%, the GAs provided by the device had an ICC similar to those calculated between the reference GA and the comparator-ultrasound as well as comparator-LMP. Moreover, the ICC of predicted GA using the established methods had exceptional values (Figure 5).

The device underestimated the reference GA 1.34 (95% CI −2.04 to −0.64) days, as well as by 0.81 (95% CI −1.50 to −0.11) days, and by 2.35 (95% CI −3.49, −1.21) days in relation to the ultrasound and LMP-GA comparators, respectively. In the meantime, the ultrasound GA comparator underestimated the reference GA by −0.53 (95% CI −0.88 to −0.19) days. The end points of the Bland-Altman 95% limits of agreement were the 2.5th percentile and 97.5th percentile for the distribution of the difference between paired measurements (Figure 5). Therefore, 95% of the differences between the new test and the reference GA were within the range of −21.2 to 18.4 days. This range was shorter than that of the comparator-LMP-GA, −25.0 to 29.0, in relation to the reference GA.

**Table 2.** Agreement between predicted gestational age and the established references.

| Statistic | Test[a] | *P* value | Reference GA[b,c] | *P* value |
|---|---|---|---|---|
| ICC[d] with reference GA (95% CI) | 0.969 (0.964 to 0.973) | N/A[e] | 1 | N/A |
| ICC with comparator-ultrasound-GA[f] (95% CI) | 0.969 (0.965 to 0.973) | N/A | 0.993 (0.992 to 0.994) | N/A |
| ICC with comparator-LMP[g]-GA[h] (95% CI) | 0.927 (0.916, 0.938) | N/A | 0.958 (0.951 to 0.964) | N/A |
| Day paired difference with reference GA (95% CI) | −1.34 (−2.04 to −0.64) | <.001 | 0 | N/A |
| Day paired difference with comparator-ultrasound-GA (95% CI) | 0−.81 (−1.50 to −.11) | <.001 | −0.53 (−0.88 to −0.19) | .002 |
| Day paired difference with LMP GA (95% CI) | −2.35 (−3.49 to −1.21) | <.001 | 0.83 (−0.07 to 1.74) | .071 |
| Bland-Altman 95% limits for the medical device (days) | N/A | N/A | −21.2 to 18.4 | N/A |
| Bland-Altman 95% limits for comparator-ultrasound (days) | −8.7 to 8.4 | N/A | −10 to 8 | N/A |
| Bland-Altman 95% limits for comparator-LMP (days) | −30 to 23.4 | N/A | −25 to 29 | N/A |

[a] Medical device gestational age predicted using the Extreme Gradient Boosting model, based on newborn skin reflectance values, birth weight, and antenatal corticosteroid therapy for fetal maturation exposure information.

[b] GA: gestational age.

[c] Reference gestational age is the best due date.

[d] ICC: intraclass correlation coefficient.

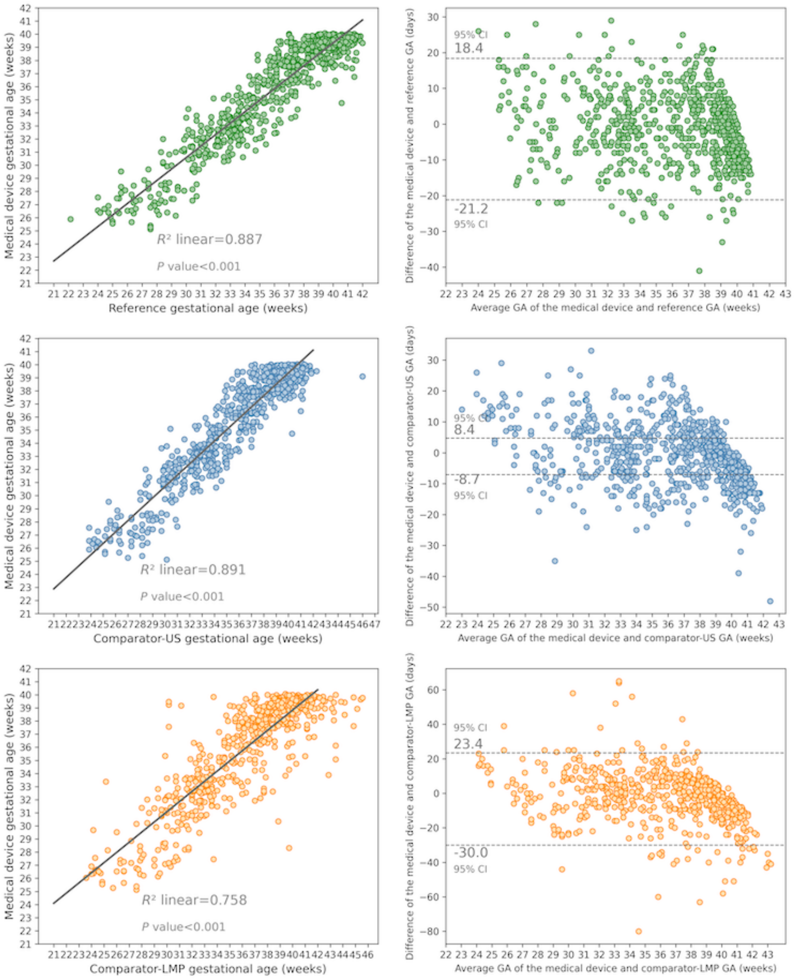
[e] N/A: not applicable.

[f] Comparator-ultrasound-GA: gestational age calculated using a second antenatal ultrasound exam after 13 weeks and 6 days of gestation and before 22 weeks.

[g] LMP: last menstrual period.

[h] Comparator-LMP-GA: the gestational age calculated using the last menstrual period.

**Figure 5.** Correlation between GAs as measured using medical devices, established methods of pregnancy dating, and Bland-Altman plots. GA: gestational age; LMP: last menstrual period; US: ultrasound.

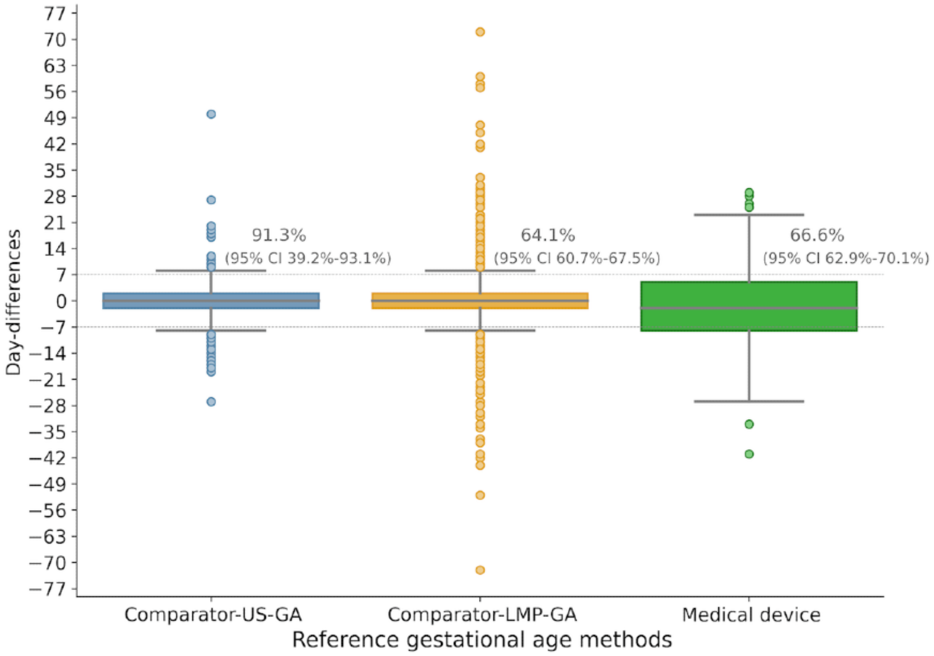


## Secondary Outcomes

### *GA Detection With 1-Week Error*

The boxplots in Figure 6 show the proportion of preterm newborns correctly detected at birth, considering an error of 1 week. We included 101 missing data points in the calculation of the rate agreement for the comparator-LMP-GA. The device achieved 66.6% (95% CI 62.9%-70.1%) of 1-week error agreement with reference pregnancy dating. This value was similar to the value of 64.1% (95% CI 60.7%-67.5%) of the comparator-LMP-GA 1-week error considering the intention-to-diagnose analysis.

**Figure 6.** Box plot of day differences between methods and reference GA, with the proportion of agreement within 7 days. GA: gestational age; LMP: last menstrual period; US: ultrasound.

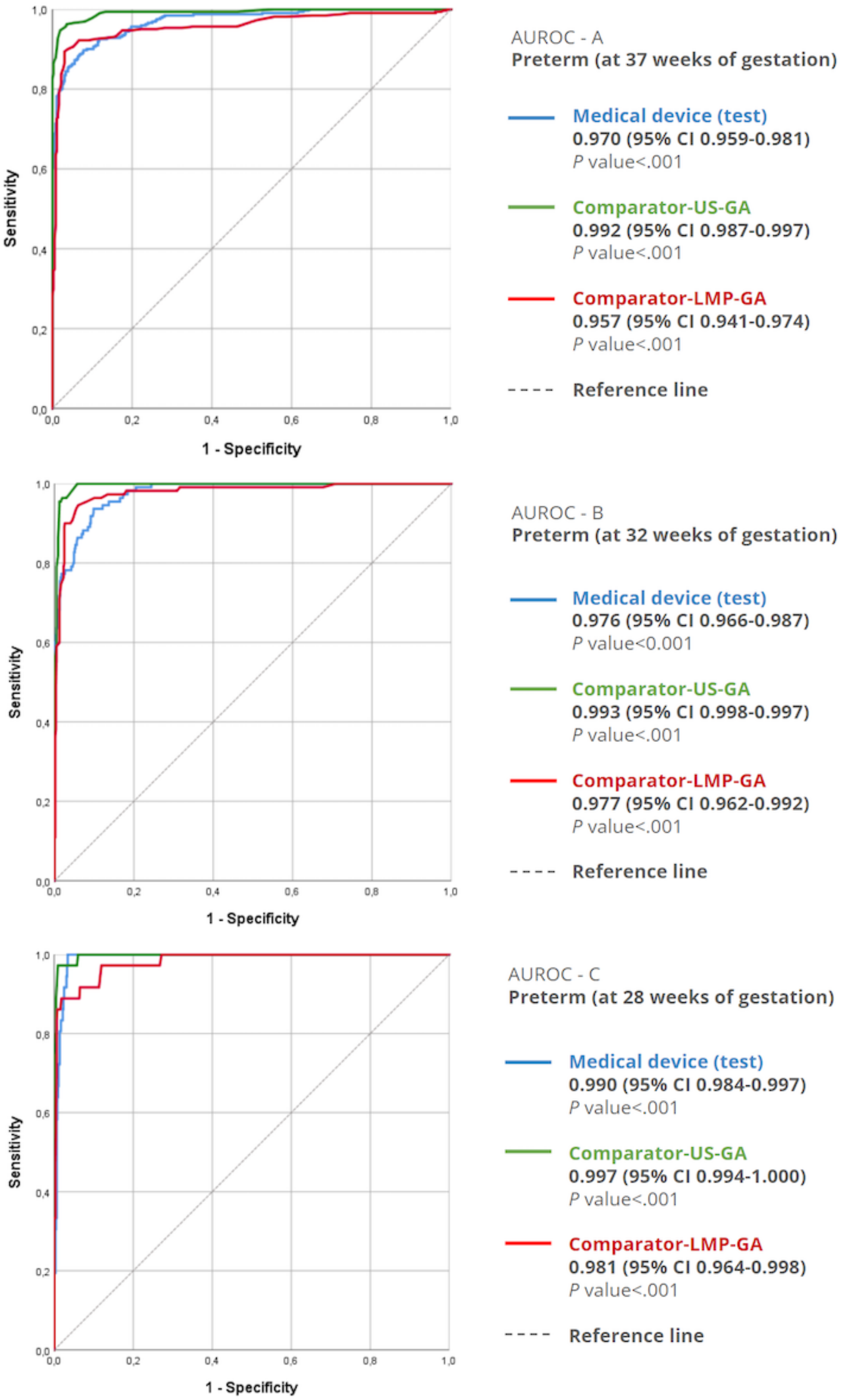


### Accuracy of the New Test for Identification of Preterm Newborns

Considering an overlap of 95% CIs in AUROC, the new test using the device had similar performance to comparator-LMP-GA in discriminating preterm against term newborns at all cutoffs, respectively, AUROC 0.973 (95% CI 0.963-0.982) and 0.957 (95% CI 0.941-0.974; Figure 7). At cutoffs after 28 and 32 weeks, the new test had similar performance compared with the comparator-ultrasound-GA.

A comprehensive analysis of the prediction accuracy for preterm newborns using the method of GA estimation and the medical device for different prematurity cutoffs is shown in Multimedia Appendix 5. Here, we draw attention to the relevant likelihood ratio, positive at 37 weeks 13.2 (95% CI 9.2-19.0) when the medical device predicts GA, showing overlaps between the comparators in terms of 95% CI 25.0 (15.4-40.4) for comparator-ultrasound-GA and 17.1 (11.0-26.6) for comparator-LMP-GA.

**Figure 7.** Receiver operating characteristic curves for the models to distinguish between term and preterm newborns. AUROC: area under the receiver operating characteristic curve; GA: gestational age; LMP: last menstrual period; US: ultrasound.



### *Intent to Perform Preterm Newborn Discrimination by the Device*

Birth care settings, where the device is to be applied, deserve an intent to perform preterm newborn discriminant analysis, simulating the existence of baseline references for GA calculation. Therefore, we considered newborns whose mothers had no recollection of LMP or unreliable information as scenario 1, corresponding to 451 (57.7%) newborns. In scenario 2, we grouped the newborns whose mothers had reliable LMP (Table 3). Concerning missing data, 3 test values for GA obtained using ACTFM machine learning imputation were valid results for this analysis. At the same time, 101 missing data items for LMP were newborns who had no comparator-LMP-GA owing to unknown menstrual dates.

The lack of a reliable LMP in scenario 1 resulted in low discrimination accuracy of 69.6% (95% CI 65.3%-73.7%) with the comparator-LMP-GA. Nevertheless, 89.6% (95% CI 86.4%-93.1%) of the newborns were correctly classified as preterm or term using the device. Great accuracy using any available method for GA estimation was observed in scenario 2, where the LMP was reliable. In such a scenario, we see similar device accuracy of 93.9% (95% CI 90.8%-96.3%) when compared with the accuracy of comparator-ultrasound-GA of 97% (95% CI 94.5%-98.5%) and comparator-LMP-GA of 93.4% (95% CI 94.5%-97.9%). The overall analysis includes crosstabs in Multimedia Appendix 6.

**Table 3.** Intent to perform preterm newborn discrimination according to simulated scenarios (N=781).

| | Scenario 1: absent or unreliable LMP[a] (n=451, 57.7%) | | | | Scenario 2: reliable LMP (n=330, 42.3%) | | | |
|---|---|---|---|---|---|---|---|---|
| | Preterm newborns | Sens[b], n/N; (95% CI) | Spec[c], n/N; (95% CI) | ACU[d], n/N; (95% CI) | Preterm newborns | Sens, n/N; (95% CI) | Spec, n/N; (95% CI) | ACU, n/N; (95% CI) |
| Reference GA[e,f] (n=781) | 199 | N/A[g] | N/A | N/A | 167 | N/A | N/A | N/A |
| Test, medical device (n=781) | 196 | 174/199; 87.4% (82%-91.7%) | 230/252; 91.3% (87.1%-94.5%) | 404/451; 89.6% (86.4%-92.2%) | 159 | 153/167; 91.6% (86.3%-95.3) | 157/163; 96.3% (92.2%-98.6%) | 310/330; 93.9% (90.8%-96.3%) |
| Comparator-ultrasound-GA[h] (n=781) | 199 | 190/199; 95.5% (91.6%-97.9%) | 241/252; 95.6% (92.3%-97.8%) | 431/451; 95.6% (93.2%-97.3%) | 167 | 162/167; 97% (93.2%-99%) | 158/163; 96.9% (93%-99%) | 320/330; 97% (94.5%-98.5%) |
| Comparator-LMP-GA[i] (n=680) | 154 | 131/199; 65.8% (59.1%-72.2%) | 183/252; 72.6% (66.9%-77.9%) | 314/451; 69.6% (65.3%-73.7%) | 167 | 160/167; 95.8% (91.6%-98.3%) | 157/163; 96.3% (93%-99%) | 317/330; 93.4% (94.5%-97.9%) |

[a] LMP: last menstrual period.

[b] Sens: sensitivity.

[c] Spec: specificity.

[d] ACU: accuracy (newborn correctly classified).

[e] GA: gestational age.

[f] Reference GA: is the best due date.

[g] N/A: not applicable.

[h] Comparator-ultrasound-GA: gestational age calculated using a second antenatal ultrasound exam after 13 weeks and 6 days of gestation and before 22 weeks.

[i] Comparator-LMP-GA: the gestational age calculated using the last menstrual period.

### *Safety of the Device*

There were no reports of unexpected medical events, unintended illness or injury, or unfortunate clinical signs in subjects, users, or others related to the investigational product. Two devices were replaced because of an unintentional drop.

## Discussion

### Principal Findings

The main contribution of this clinical trial is the validation of a new approach for GA estimation, regardless of fetal ultrasound measures by demonstrating accurate outcomes. Based on birth weight, ACTFM exposure data, and use of a handled medical device to assess skin maturity and process algorithms, 91.4% (714/781) of newborns were correctly classified. A reliable antenatal age is a prerequisite for preterm newborn classification in birth care settings and is the first step in delivering the necessary care, considering the risks of prematurity. A term newborn, together with good tonus, breathing, or crying, is an essential element to determine steps of newborn resuscitation [30]. Although that statement seems very simple, it is quite far from reality. Without certainty as to the day in the female cycle on which conception occurred, ultrasound measurement of the crown-rump length is a consensual reference for redating pregnancy estimated by the LMP [8]. This dependence on early echographic scans has deprived many pregnant women and their babies of trustable GA [10]. Such a technological gap causes even more disparities than the difference between childbirth scenarios in fully equipped facilities and those ill-equipped with scarce technology. Moreover, it can impair the correct classification of infants as premature or growth restricted [31]. Whereas the underestimation of GA by 1.34 days on average in our results could reverberate in over care of a newborn with device implementation, neglecting a newborn at risk owing to the lack of GA data is still the worst. We believe that the risks attributed to preterm infants and the benefit of early diagnosis outweigh overdiagnosis. In addition, the delivery of neonatal care at birth is based on a set of clinical parameters, including GA [32].

In this combined study covering the enhancement of the prediction model for postnatal GA and validation of the device, we believe that the application of k-fold cross-validation with the use of machine learning algorithms provided accurate predictions [33]. While large data samples are unavailable, the process of training and testing was able to estimate the performance of algorithms until we have finished other ongoing clinical trials for external validation [34]. Furthermore, the quantification of uncertainty intervals regarding the predicted GA (calculated in days) and comparisons with established references allowed the simulation of realistic scenarios for application. Besides, the CIs accompanying AUROC accuracy contributed to revealing the forecast's limits for discriminating terms from preterm newborns at different cutoff points with clinical relevance. Such strengths are critical for ensuring the potential value of the device in facing the challenges of postnatal identification of preterm newborns [35]. Postnatal approaches for GA assessment had characteristically shown higher errors than antenatal approaches [36]; however, studies using first-trimester ultrasound as the standard for postnatal GA comparisons were uncommon until recently. In a recent study comparing the accuracy of postnatal GA assessment, the New Ballard Score obtained −2.93 to 2.65 weeks of error compared with early ultrasound reference, analyzing a sample with 78.3% of preterm newborns [37]. In our study, the limits were −21.2 to 18.4 days, even though we did not compare the results from the medical device with any postnatal reference, it was a promising result.

Thus, data science algorithms have thus emerged with the aim of qualifying pregnancy dating. High-performance reports using learning models based on antenatal ultrasound predictors [38] contradistinguished meager outcomes from those using other morphometric postnatal predictors [3]. Moreover, valuable algorithms with postnatal combinations on the maturity scores of newborns are promising, even demanding special skills to apply [12]. Underqualified birth attendants represent a challenge in developing countries, further limiting the use of existing birth care solutions [39]. One advantage of our device is the skin assessment automation that notifies measurement errors caused by the movement of the newborn or examiner. Previous reports have detailed the human skin's light-skin interaction and optical properties that benefit this technology [18,25].

The device's predictive XGBoost algorithm used information that health professionals could quickly obtain in childbirth settings—the birth weight and the ACTFM exposure—and that could add value to the visual appearance of skin maturity. Explaining the model used during development, we have already demonstrated that the multivariate model for predicting GA, combining the skin reflection with birth weight, was better than these variables apart [18]. In this clinical trial, the choice had biological plausibility extending beyond mathematical reasons. Birth weight assessment is a universal step of primary routines in childbirth settings [6]. Meanwhile, predicting preterm birth based on birth weight when lacking a gold standard is far from a perfect solution. There is prior scientific evidence that birth weight is not sufficient to predict GA or a preterm newborn [9]. The weight at birth results from the dynamic process of past intrauterine growth beyond the gestation length [11]. Otherwise, the physical and neurological characteristics of maturity of the newborn adding value to predict GA are already extensively used and validated in the postnatal scores [36]. Meanwhile, the postnatal scores of newborn maturity, as the only method, have shown low accuracy in determining GA and identifying prematurity [36]. We combined birth weight and skin maturity adjusted to the ACTFM to predict GA, representing the clinical rationale with high $R^2$ and low MAE, thereby avoiding the standalone model with birth weight (Multimedia Appendix 4).

In this trial, the GA estimated through using the device had great agreement with the reference GA at birth. The Bland-Altman test (95% limit) was lower than the comparator-LMP-GA. Moreover, this device could provide a GA to handle situations without ACTFM information as a potential tool in low-resource birth settings. Considering the simulated scenario with LMP either absent or unreliable (n=451 newborns), the new test had a better performance than the comparator-LMP for the estimation of GA. This result highlighted the context of use of this medical device, as the GA based on memory recall of the LMP missed 68 out of 199 preterm newborns, expressing a lower sensitivity when we applied the intent-to-discriminate analysis.

## Strengths and Limitations

Exposure to ACTFM played an uncertain role in the predictive model. Nevertheless, there was a rationale to consider its importance to adjust the skin reflection. Antenatal corticosteroids to improve newborn outcomes are an evidence-based intervention recommended for women at risk of preterm birth [32]. However, in addition to the acceleration of lung maturity, the effect of the drug occurs in other organs. The early fetal presence of receptors of corticosteroid hormone receptors in skin epithelial cells indicates that glucocorticoids may play an important role in the differentiation and development of human skin [40]. However, clinical evidence of the effect of ACTFM exposure on skin maturity remains unsubstantiated [41]. Thus, the adoption of the new test warrants caution. Thus, until proven otherwise, we consider that the importance of ACTFM exposure data to adjust the GA modeling is related to an effect on skin maturity. Even so, we cannot deny that antenatal exposure to corticoid therapy is more common in premature infants—264 (72.3%) of the preterm newborns in this study. In this respect, this regressor variable could imply a bias toward preterm newborn detection. The aforementioned ongoing study for external validation of the algorithms could further elucidate this issue because the enrollment process of newborns introduced the Mozambican birth scenario, where, unfortunately, ACTFM is not guaranteed for every woman at risk of preterm birth [34]. Furthermore, the accuracy is unknown for newborns with diseases that alter skin structure, which is an exclusion criterion in this study.

Current approaches to calculating GA are sensitive to data quality, resulting in an inappropriate classification of prematurity [9]. This study was committed to representing a realistic scenario regarding data quality, as stated in the research protocol, with data collection and curation to ensure the best reference and comparators for analysis. Before opening the blinding of the trial, a consistent process confronted data entries with digital images of the clinical documents taken during enrollment. Furthermore, dedicated software was developed exclusively for clinical trials, considering the quality and constraints of the variables. Part of the enrollment occurred during the COVID-19 pandemic, resulting in a minimal amount of missing data, such as yes or no for ACTFM) information (3/781, 0.4% of newborns). The lack of an LMP reference, antenatal care without early antenatal ultrasound, and unqualified date recollection for pregnancy dating at birth justify efforts to enhance the reliability of pregnancy dating through more accurate and accessible technologies to improve pregnancy outcomes and neonatal survival [10]. In our study, qualifying the LMP at birth with questions about memory regarding dates and menstrual cycles, and checking antenatal clinical documents at birth provided an estimation of GA to identify 160 preterm newborns among 167, when available.

Regarding the generalizability of the outcomes, this multicenter trial gathered referral perinatal units from Brazil's northern, central, southwestern, and southern regions. This collaborative evaluation contributed to obtaining a sample of a mixed population of newborns with high miscegenation and involved 15 examiners who attended good clinical practice training. Both intraobserver and interobserver errors of the measurements were low, in line with previous results [18]. The number of preterm newborns was sufficient to analyze subcategories of prematurity as extreme preterm (n=42); however, the overall rate of preterm newborns was 46.9%, values observed in referral facilities for high-complexity perinatal care and not in the general population of Brazilian newborns [42]. Thus, such a high frequency might limit the representativeness of the results for the general population of newborns in low-complexity settings, where the prematurity rate is approximately 11% [1]. Among the 781 newborns, neonatal deaths during 72 hours of follow-up occurred in 14 (1.8%), with 12 deaths occurring in newborns with GA <28 weeks owing to complications arising from extreme prematurity. We expect to target the worst childbirth scenarios for this technology implementation [39]. In addition, the safety of this device is similar to that of other optical technologies already used in neonatal care [30].

## Conclusions

The assessment of newborn's skin maturity adjusted by learning models promises accurate pregnancy dating at birth, even without the antenatal ultrasound reference. Identifying preterm newborns is the first step toward meeting their needs. The global rate of neonatal mortality is approximately 6700 neonatal deaths daily, mostly from preventable or treatable conditions in scenarios without adequate health care [43]. Without proper comparisons, the device had a lower error range than after-birth maturity scores. To provide future evidence, comparisons are expected based on postnatal approaches for GA estimation, such as scores of maturity and foot length, or image combinations [3]. We hope that strengthening the data sources of health care facilities with a reliable GA can help identify vulnerable newborns in situations without such information.

## Acknowledgments

The authors would like to thank Grand Challenges Exploration from the Bill & Melinda Gates Foundation, who supported this research (grant OPP1128907, contract), and *Fundação de Amparo a Pesquisa de Minas Gerais*, Brazil, nonprofit sectors. This clinical trial was funded by the Brazilian Ministry of Health, Program of Development of the Industrial Health Complex, project 23072.052747/2017-51; Grand Challenges Exploration from the Bill & Melinda Gates Foundation; and *Fundação de Amparo a Pesquisa de Minas Gerais*, Brazilian Ministry of Health, CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico Nº 03/2021. ZSNR is a researcher with a grant from the *Conselho Nacional de Pesquisa* (CNPq 305837/2021-4), Brazil. The final formatting of the English review manuscript was supported by Fiocruz-Brazil and the Grand Challenges, Canada. The authors would like to thank all the clinical staff at the participating hospitals who cared for the patients and undergraduate students who helped with data collection.

## Data Availability

Data are available upon reasonable request and after anonymization to allow for the sharing of data ethically and legally, thus preserving the confidentiality of the persons who participated in this study. The study protocol, statistical analysis plan, and informed consent form are available in previous publications. Standard operational procedures with detailed methods have been deposited in the protocols.io website.

## Authors' Contributions

The lead authors (ZSNR, RMdCR, RNG, JdSG, and RAPLdA) had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. Correspondence and requests for materials should be addressed to ZSNR.

## Conflicts of Interest

The authors declare a patent deposit on behalf of the Universidade Federal de Minas Gerais and Fundação de Amparo a Pesquisa de Minas Gerais, Brazil. The inventors were ZSNR, RNG, and BR1020170235688 (CTIT-PN862). BirthTech, a spin-off company, received a license to produce and commercialize this technology, and RNG is its founder. No sponsor had any role in the study design, data collection, data analysis, data interpretation, writing, or decision to submit the manuscript.

Multimedia Appendix 1
Database of clinical variables collected from each newborn and Preemie-Test skin acquisitions.
[DOCX File , 17 KB - jmir_v24i9e38727_app1.docx ]

---

Multimedia Appendix 2
The reliability of the skin assessment with the photometer of the device.
[DOCX File , 12 KB - jmir_v24i9e38727_app2.docx ]

---

Multimedia Appendix 3
Analytical pipeline.
[DOCX File , 60 KB - jmir_v24i9e38727_app3.docx ]

---

Multimedia Appendix 4
Correlation between reference gestational age and predictor variables.
[DOCX File , 23 KB - jmir_v24i9e38727_app4.docx ]

---

Multimedia Appendix 5
Accuracy for preterm newborn discrimination according to the methods of gestational age estimation.
[DOCX File , 14 KB - jmir_v24i9e38727_app5.docx ]

---

Multimedia Appendix 6
Intent to perform preterm newborn discrimination according to simulated scenarios of care.
[DOCX File , 17 KB - jmir_v24i9e38727_app6.docx ]

---

## References

1. Walani SR. Global burden of preterm birth. Int J Gynaecol Obstet 2020 Jul;150(1):31-33. [doi: 10.1002/ijgo.13195] [Medline: 32524596]
2. Oza S, Cousens SN, Lawn JE. Estimation of daily risk of neonatal death, including the day of birth, in 186 countries in 2013: a vital-registration and modelling-based study. Lancet Glob Health 2014 Nov;2(11):e635-e644 [FREE Full text] [doi: 10.1016/S2214-109X(14)70309-2] [Medline: 25442688]
3. Torres M, Valstar MF, Henry C, Ward C, Sharkey D. Small sample deep learning for newborn gestational age estimation. In: Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition. 2017 Presented at: FG '17; May 30-June 3, 2017; Washington, DC, USA p. 79-86. [doi: 10.1109/fg.2017.19]
4. Guo X, Li X, Qi T, Pan Z, Zhu X, Wang H, et al. A birth population-based survey of preterm morbidity and mortality by gestational age. BMC Pregnancy Childbirth 2021 Apr 10;21(1):291 [FREE Full text] [doi: 10.1186/s12884-021-03726-4] [Medline: 33838659]
5. Kim ET, Singh K, Moran A, Armbruster D, Kozuki N. Obstetric ultrasound use in low and middle income countries: a narrative review. Reprod Health 2018 Jul 20;15(1):129 [FREE Full text] [doi: 10.1186/s12978-018-0571-y] [Medline: 30029609]
6. Tunçalp Ö, Pena-Rosas JP, Lawrie T, Bucagu M, Oladapo OT, Portela A, et al. WHO recommendations on antenatal care for a positive pregnancy experience-going beyond survival. BJOG 2017 May;124(6):860-862. [doi: 10.1111/1471-0528.14599] [Medline: 28190290]
7. Lawn JE, Blencowe H, Oza S, You D, Lee AC, Waiswa P, Lancet Every Newborn Study Group. Every Newborn: progress, priorities, and potential beyond survival. Lancet 2014 Jul 12;384(9938):189-205. [doi: 10.1016/S0140-6736(14)60496-7] [Medline: 24853593]
8. Committee on Obstetric Practice American Institute of Ultrasound in Medicine Society for Maternal–Fetal Medicine. Committee opinion no 700: methods for estimating the due date. Obstet Gynecol 2017 May;129(5):e150-e154. [doi: 10.1097/AOG.0000000000002046] [Medline: 28426621]
9. Miller L, Wanduru P, Santos N, Butrick E, Waiswa P, Otieno P, et al. Working with what you have: how the East Africa Preterm Birth Initiative used gestational age data from facility maternity registers. PLoS One 2020 Aug 31;15(8):e0237656 [FREE Full text] [doi: 10.1371/journal.pone.0237656] [Medline: 32866167]
10. Karl S, Li Wai Suen CS, Unger HW, Ome-Kaius M, Mola G, White L, et al. Preterm or not--an evaluation of estimates of gestational age in a cohort of women from Rural Papua New Guinea. PLoS One 2015 Aug 6;10(5):e0124286 [FREE Full text] [doi: 10.1371/journal.pone.0124286] [Medline: 25945927]
11. Damhuis SE, Ganzevoort W, Gordijn SJ. Abnormal fetal growth: small for gestational age, fetal growth restriction, large for gestational age: definitions and epidemiology. Obstet Gynecol Clin North Am 2021 Jun;48(2):267-279 [FREE Full text] [doi: 10.1016/j.ogc.2021.02.002] [Medline: 33972065]

12. Rittenhouse KJ, Vwalika B, Keil A, Winston J, Stoner M, Price JT, et al. Improving preterm newborn identification in low-resource settings with machine learning. PLoS One 2019 Feb 27;14(2):e0198919 [FREE Full text] [doi: 10.1371/journal.pone.0198919] [Medline: 30811399]
13. Kullinger M, Granfors M, Kieler H, Skalkidou A. Discrepancy between pregnancy dating methods affects obstetric and neonatal outcomes: a population-based register cohort study. Sci Rep 2018 May 02;8(1):6936 [FREE Full text] [doi: 10.1038/s41598-018-24894-y] [Medline: 29720591]
14. Global strategy on digital health 2020-2025. World Health Organization. Geneva, Switzerland: World Health Organization; 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2022-06-24]
15. Nelson GA, Holschuh C. Evaluation of telehealth use in prenatal care for patient and provider satisfaction: a step toward reducing barriers to care. J Nurse Pract 2021 Apr 1;17(4):481-484. [doi: 10.1016/j.nurpra.2020.12.026]
16. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017 Sep 14;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
17. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? Am J Med 2018 Feb;131(2):129-133. [doi: 10.1016/j.amjmed.2017.10.035] [Medline: 29126825]
18. Reis ZS, Vitral GL, de Souza IM, Rego MA, Guimaraes RN. Newborn skin reflection: proof of concept for a new approach for predicting gestational age at birth. A cross-sectional study. PLoS One 2017 Sep 20;12(9):e0184734 [FREE Full text] [doi: 10.1371/journal.pone.0184734] [Medline: 28931040]
19. 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]
20. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
21. Venema E, Wessler BS, Paulus JK, Salah R, Raman G, Leung LY, et al. Large-scale validation of the prediction model risk of bias assessment Tool (PROBAST) using a short form: high risk of bias models show poorer discrimination. J Clin Epidemiol 2021 Oct;138:32-39 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.06.017] [Medline: 34175377]
22. Reis ZS, Guimarães RN, Rego MA, Maia de Castro Romanelli R, Gaspar JD, Vitral GL, et al. Prematurity detection evaluating interaction between the skin of the newborn and light: protocol for the preemie-test multicentre clinical trial in Brazilian hospitals to validate a new medical device. BMJ Open 2019 Mar 05;9(3):e027442 [FREE Full text] [doi: 10.1136/bmjopen-2018-027442] [Medline: 30842119]
23. Reis Z, Vitral GL, Guimarães RN, de Aguiar RA, Romanelli RM. The Preemie-Test for the assessment of the newborn skin maturity. Universidade Federal de Minas Gerais. 2019 Oct. URL: https://www.protocols.io/view/the-preemie-test-for-the-assessment-of-the-newborn-rm7vz8njxvx1/v1 [accessed 2021-09-11]
24. Reis Z, Gaspar J, Elias SO, de Aguiar RA. Algorithm for gestational age assessment at birth. Universidade Federal de Minas Gerais. 2020 Jan 17. URL: https://www.protocols.io/view/algorithm-for-gestational-age-assessment-at-birth-eq2ly3j9pgx9/v1 [accessed 2021-09-11]
25. Silva PC, Guimarães RN, Souza RG, Reis ZS. A quantitative cross-sectional analysis of the melanin index in the skin of preterm newborns and its association with gestational age at birth. Skin Res Technol 2020 May;26(3):356-361 [FREE Full text] [doi: 10.1111/srt.12810] [Medline: 31763716]
26. Popović ZB, Thomas JD. Assessing observer variability: a user's guide. Cardiovasc Diagn Ther 2017 Jun;7(3):317-324 [FREE Full text] [doi: 10.21037/cdt.2017.03.12] [Medline: 28567357]
27. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16; August 13-17, 2016; San Francisco, CA, USA p. 785-794. [doi: 10.1145/2939672.2939785]
28. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2. 1995 Presented at: IJCAI '95; August 20-25, 1995; Montreal, Canada p. 1137-1143.
29. Friedman JH. Stochastic gradient boosting. Comput Stat Data Analysis 2002 Feb 28;38(4):367-378. [doi: 10.1016/s0167-9473(01)00065-2]
30. Wyckoff MH, Wyllie J, Aziz K, de Almeida MF, Fabres JW, Fawke J, Neonatal Life Support Collaborators. Neonatal life support: 2020 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Circulation 2020 Oct 20;142(16_suppl_1):S185-S221 [FREE Full text] [doi: 10.1161/CIR.0000000000000895] [Medline: 33084392]
31. Scott K, Gupta S, Williams E, Arthur M, Somayajulu UV, Noguchi L. "I can guess the month … but beyond that, I can't tell" an exploratory qualitative study of health care provider perspectives on gestational age estimation in Rajasthan, India. BMC Pregnancy Childbirth 2020 Sep 11;20(1):529 [FREE Full text] [doi: 10.1186/s12884-020-03201-6] [Medline: 32917163]
32. Emeruwa UN, Krenitsky NM, Sheen JJ. Advances in management for preterm fetuses at risk of delivery. Clin Perinatol 2020 Dec;47(4):685-703. [doi: 10.1016/j.clp.2020.08.006] [Medline: 33153655]

33. Weiss SM, Indurkhya N. Rule-based machine learning methods for functional prediction. J Artif Intell Res 1995 Dec 01;3:383-403. [doi: 10.1613/jair.199]
34. Reis Z, Vitral G, Guimarães R, Gaspar J, Colosimo E, Taunde S, et al. Premature or small for gestational age discrimination: international multicenter trial protocol for classification of the low-birth-weight newborn through the optical properties of the skin. JMIR Res Protoc 2020 Jul 14;9(7):e16477 [FREE Full text] [doi: 10.2196/16477] [Medline: 32673275]
35. Ananth CV, Brandt JS. Fetal growth and gestational age prediction by machine learning. Lancet Digit Health 2020 Jul;2(7):e336-e337 [FREE Full text] [doi: 10.1016/S2589-7500(20)30143-6] [Medline: 33328092]
36. Lee AC, Panchal P, Folger L, Whelan H, Whelan R, Rosner B, et al. Diagnostic accuracy of neonatal assessment for gestational age determination: a systematic review. Pediatrics 2017 Dec;140(6):e20171423. [doi: 10.1542/peds.2017-1423] [Medline: 29150458]
37. Stevenson A, Joolay Y, Levetan C, Price C, Tooke L. A comparison of the accuracy of various methods of postnatal gestational age estimation; including ballard score, foot length, vascularity of the anterior lens, last menstrual period and also a clinician's non-structured assessment. J Trop Pediatr 2021 Jan 29;67(1):fmaa113. [doi: 10.1093/tropej/fmaa113] [Medline: 33693940]
38. Fung R, Villar J, Dashti A, Ismail LC, Staines-Urias E, Ohuma EO, International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). Achieving accurate estimates of fetal gestational age and personalised predictions of fetal growth based on data from an international prospective cohort study: a population-based machine learning study. Lancet Digit Health 2020 Jul;2(7):e368-e375 [FREE Full text] [doi: 10.1016/S2589-7500(20)30131-X] [Medline: 32617525]
39. Soubeiga D, Gauvin L, Hatem MA, Johri M. Birth Preparedness and Complication Readiness (BPCR) interventions to reduce maternal and neonatal mortality in developing countries: systematic review and meta-analysis. BMC Pregnancy Childbirth 2014 Apr 04;14:129 [FREE Full text] [doi: 10.1186/1471-2393-14-129] [Medline: 24708719]
40. Condon J, Gosden C, Gardener D, Nickson P, Hewison M, Howie AJ, et al. Expression of type 2 11beta-hydroxysteroid dehydrogenase and corticosteroid hormone receptors in early human fetal life. J Clin Endocrinol Metab 1998 Dec;83(12):4490-4497. [doi: 10.1210/jcem.83.12.5302] [Medline: 9851798]
41. August D, Kandasamy Y. The effects of antenatal glucocorticoid exposure on fetal and neonatal skin maturation. J Perinat Med 2017 Nov 27;45(8):969-975. [doi: 10.1515/jpm-2016-0338] [Medline: 28170341]
42. do Carmo Leal M, Szwarcwald CL, Almeida PV, Aquino EM, Barreto ML, Barros F, et al. Saúde reprodutiva, materna, neonatal e infantil nos 30 anos do Sistema Único de Saúde (SUS). Cien Saude Colet 2018 Jun;23(6):1915-1928 [FREE Full text] [doi: 10.1590/1413-81232018236.03942018] [Medline: 29972499]
43. IGME: Levels and trends in child mortality. United Nations International Children's Emergency Fund. 2020. URL: https://data.unicef.org/resources/levels-and-trends-in-child-mortality/ [accessed 2021-09-11]

## Abbreviations

**ACTFM:** antenatal corticosteroid therapy for fetal maturation
**AUROC:** area under the receiver operating characteristic curve
**GA:** gestational age
**ICC:** intraclass correlation coefficient
**LMP:** last menstrual period
**MAE:** mean absolute error
**XGBoost:** Extreme Gradient Boosting

*Edited by R Kukafka; submitted 13.04.22; peer-reviewed by MS Arefin, S Ouyang, P von Wedel; comments to author 11.06.22; revised version received 04.07.22; accepted 25.07.22; published 07.09.22.*

*Please cite as:*
*Reis ZSN, Romanelli RMDC, Guimarães RN, Gaspar JDS, Neves GS, do Vale MS, Nader PDJ, de Moura MDR, Vitral GLN, dos Reis MAA, Pereira MMM, Marques PF, Nader SS, Harff AL, Beleza LDO, de Castro MEC, Souza RG, Pappa GL, de Aguiar RAPL*
*Newborn Skin Maturity Medical Device Validation for Gestational Age Prediction: Clinical Trial*
*J Med Internet Res 2022;24(9):e38727*
*URL: https://www.jmir.org/2022/9/e38727*
*doi:10.2196/38727*
*PMID:36069805*

Original Paper

# Real-world Implementation of an eHealth System Based on Artificial Intelligence Designed to Predict and Reduce Emergency Department Visits by Older Adults: Pragmatic Trial

Joël Belmin[1,2], MD, PhD; Patrick Villani[3,4], MD, PhD; Mathias Gay[5], PhD; Stéphane Fabries[6], MS; Charlotte Havreng-Théry[2,7], PhD; Stéphanie Malvoisin[8], MD; Fabrice Denis[9], MD, PhD; Jacques-Henri Veyron[7], MS

[1]Hôpital Charles Foix, Assistance Publique-Hôpitaux de Paris, Ivry-sur-Seine, France

[2]Laboratoire Informatique Médicale et Ingénierie des Connaissances en eSanté (UMRS 1142), Institut National de la Santé et de la Recherche Médicale and Sorbonne Université, Paris, France

[3]Unité de Médecine Interne, Gériatrie et Thérapeutique, Assistance Publique-Hôpitaux de Marseille, Marseille, France

[4]Etablissement Français du Sang, Anthropologie bio-culturelle, Droit, Ethique et Santé, Centre National de la Recherche Scientifique, Université Aix-Marseille, Marseille, France

[5]Communauté professionnelle de santé Itinéraire Santé, Marseille, France

[6]Intervenants Libéraux et Hospitaliers Unis pour le Patient, Marseille, France

[7]Presage, Paris, France

[8]Centre hospitalo-universitaire La Réunion, Saint-Pierre, French Southern Territories

[9]Institut Inter-Régional de Cancérologie Jean Bernard, Le Mans, France

**Corresponding Author:**
Jacques-Henri Veyron, MS
Presage
72 boulevard de Sébastopol
Paris, 75003
France
Phone: 33 149964242
Email: jhveyron@presage.care

## Abstract

**Background:** Frail older people use emergency services extensively, and digital systems that monitor health remotely could be useful in reducing these visits by earlier detection of worsening health conditions.

**Objective:** We aimed to implement a system that produces alerts when the machine learning algorithm identifies a short-term risk for an emergency department (ED) visit and examine health interventions delivered after these alerts and users' experience. This study highlights the feasibility of the general system and its performance in reducing ED visits. It also evaluates the accuracy of alerts' prediction.

**Methods:** An uncontrolled multicenter trial was conducted in community-dwelling older adults receiving assistance from home aides (HAs). We implemented an eHealth system that produces an alert for a high risk of ED visits. After each home visit, the HAs completed a questionnaire on participants' functional status, using a smartphone app, and the information was processed in real time by a previously developed machine learning algorithm that identifies patients at risk of an ED visit within 14 days. In case of risk, the eHealth system alerted a coordinating nurse who could then inform the family carer and the patient's nurses or general practitioner. The primary outcomes were the rate of ED visits and the number of deaths after alert-triggered health interventions (ATHIs) and users' experience with the eHealth system; the secondary outcome was the accuracy of the eHealth system in predicting ED visits.

**Results:** We included 206 patients (mean age 85, SD 8 years; 161/206, 78% women) who received aid from 109 HAs, and the mean follow-up period was 10 months. The HAs monitored 2656 visits, which resulted in 405 alerts. Two ED visits were recorded following 131 alerts with an ATHI (2/131, 1.5%), whereas 36 ED visits were recorded following 274 alerts that did not result in an ATHI (36/274, 13.4%), corresponding to an odds ratio of 0.10 (95% IC 0.02-0.43; *P*<.001). Five patients died during the study. All had alerts, 4 did not have an ATHI and were hospitalized, and 1 had an ATHI (*P*=.04). In terms of overall usability, the digital system was easy to use for 90% (98/109) of HAs, and response time was acceptable for 89% (98/109) of them.

**Conclusions:** The eHealth system has been successfully implemented, was appreciated by users, and produced relevant alerts. ATHIs were associated with a lower rate of ED visits, suggesting that the eHealth system might be effective in lowering the number of ED visits in this population.

**Trial Registration:** clinicaltrials.gov NCT05221697; https://clinicaltrials.gov/ct2/show/NCT05221697.

***(J Med Internet Res 2022;24(9):e40387)*** doi:10.2196/40387

**KEYWORDS**

emergency department visits; home care aides; community-dwelling older adults; smartphone; mobile phone; predictive tool; health intervention; machine learning; predict; risk; algorithm; model; user experience; alert; monitoring

## Introduction

The aging human population is increasing worldwide, and their health is characterized by high prevalence of chronic diseases and multimorbidity and a high vulnerability to acute diseases [1,2]. A large proportion of older adults go through emergency department (ED) visits and unplanned hospitalizations, and this proportion increases with advancing age and frailty [3]. In the United States, almost one of every three US emergency medical services responses involves an older adult [4]. In 80% of cases, an older adult's ED visit is followed by an unscheduled hospitalization, and therefore, has a high medical and economic cost [4]. ED visits and hospitalizations can have a negative impact on the health status of frail older patients by decreasing their functional capacities, which may persist for a long time thereafter [5,6]. Since a large proportion of ED visits are avoidable (range 8%-62%) [7-10], strategies to identify high-risk patients and enable them to be treated in outpatient care settings might help improve the appropriate use of ED visits and control health expenditures [11].

Patient (or family)-reported outcome measure (PROM) systems benefit patients with chronic diseases by improving quality of life, reducing mortality, reducing ED visits, and hospitalizations [12-14]. In 2019, we conducted an observational cohort study, enrolling 301 older individuals who received regular visits by home aides (HAs); we developed a machine learning algorithm to predict the risk of emergency visits, with a prediction window of 7-14 days and a predictive performance (ie, the area under the receiver operating characteristic curve) of 0.70 after 7 days and 0.67 after 14 days [15]. This algorithm opens the possibility of mobilizing health professionals to intervene early in an acute illness or in the decompensation of a chronic illness before they lead to an ED visit and unplanned hospitalization. This represents a significant advance over existing scores with a predictive window of 6-24 months [16-18], which can lead to preventive actions that are temporally distant from events that lead to emergencies.

Today, attending physicians or nurses no longer have time and opportunity to regularly visit older people at home. HAs are key professionals in maintaining older adults at home. They have regular contact with them and can provide important information for decompensation prevention. The idea of this system is to optimize medical interventions when they are really necessary and to find alternatives via health recommendations or other interventions that do not require attending physicians when their presence is dispensable. It allows us to value HAs' job, and it is based on their proximity with older adults to optimize the care pathway and avoid ED visits. Very few studies in the literature have analyzed the effectiveness of community-based interventions to prevent avoidable emergency hospitalizations of older individuals [16,17]. Recently, Nord et al [18] obtained a 17% decrease in hospitalization rate of older adults in primary care settings, by providing a nurse visit based on comprehensive geriatric assessment among older adults considered at risk for an ED visit by a 12-month predictive tool.

We have conducted a real-world pragmatic trial that included older adults receiving assistance from HAs. We aimed to implement a system that produces alerts when the machine learning algorithm identifies a short-term risk for an ED visit and examine health interventions required after these alerts and users' experience. This study highlights the feasibility of the general system, the levers for compliance improvement, and its optimal effectiveness in reducing emergency hospitalizations among older people living at home.

## Methods

### Study Design and Recruitment

This multicenter uncontrolled pragmatic trial (NCT05221697) was conducted with 3 home aid facilities participating in the study, located in 3 French cities: Marseille, Versailles, and Dinan. To be eligible, participants should be aged ≥75 years and living at home, receive the help of a social worker from these facilities, have seen their general practitioner within the last 12 months and had a mild or moderate level of dependency according to French national dependency tool, the AGGIR scale [19]. Written consent was obtained before inclusion in the study. Screening and enrollment started on July 1, 2020, and data were collected from September 1, 2020, to August 31, 2021. Participants' family caregivers and general practitioners received information about the participation of their proxy or patient in the study, as well as the patient's nurse if the patient received nursing care at home. Participants' demographics, housing, family situation, dependency level, hospitalization (dates), and death information were collected by the managers of the home aid facilities.

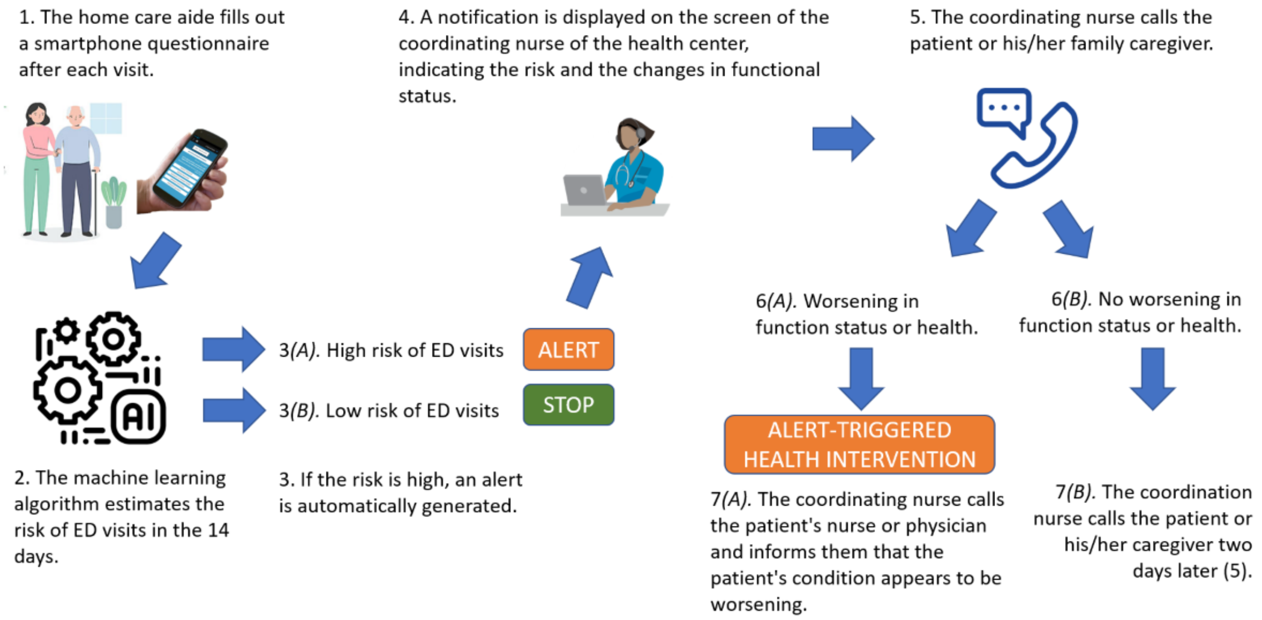
### Intervention

The intervention is summarized in Figure 1. HAs of these facilities were equipped with a smartphone app and were provided with a user manual, defining the app functioning. HAs were asked to complete a simple questionnaire after each home visit, via the smartphone app (Multimedia Appendix 1), which included a user manual, defining the different items in the questionnaire and the answer options.

This questionnaire focused on functional and clinical autonomy (ie, activities of daily life), possible medical symptoms (eg, fatigue, falls, and pain), changes in behavior (eg, recognition and aggressiveness), and communication with the HA or their surroundings. This questionnaire is composed of very simple and easy-to-understand questions, giving a global view of the person's condition. For each of the 23 questions, a yes/no answer was requested. Data recorded by HAs were sent in real time to a secure server to be analyzed by our machine learning algorithm, which predicted the risk level and displayed it on a web-based secure medical device called PRESAGE CARE, which is CE marked. A simplified diagram of the processing system and the description of data processing system can be found in Multimedia Appendix 2. Particularly, when the algorithm predicted a high-risk level, an alert was displayed in the form of a notification on the screen to the coordinating nurse of the health care network center of the district (Dispositif d'Appui à la Coordination of the Agence Régionale de Santé). This risk notification was accompanied by information about recent changes in the patients' functional status, identified from the HAs' records, to assist the coordinating nurse in interacting with family caregiver and other health professionals.

In the event of an alert, the coordinating nurse called the family caregiver to inquire about recent changes in the patient's health condition and for doubt removal and could then decide to ask for a health intervention according to a health intervention model developed before the start of the study. In brief, this alert-triggered health intervention (ATHI) consisted of calling the patient's nurse (if the patient had regular home visits of a nurse) or the patient's general practitioner and informing them of a worsening of the patient's functional status and a potential risk of an ED visit in the next few days according to the eHealth system algorithm. The ATHI was performed with the natural resources of the health system and not with the physicians or nurses employed in the study. No specific instruction or protocol was given to these health professionals, and they were free to make their own decisions. This model of ATHI had been presented and approved by the Agences Régionales de Santé of the regions involved in our study (Figure 1).

**Figure 1.** The application of the intervention protocol for alert management. ED: emergency department.



## Outcomes

The primary outcomes were the rate of ED visits and the number of deaths recorded by the coordinating nurses, as well as the users' experience with HAs and coordinating nurses recorded by the questionnaires. The eHealth system's organizational outcomes in accordance with the guidelines for the evaluation of eHealth systems of the Haute Autorité de Santé, the French national health agency [20], were as follows: the number of monitored visits, defined as home visits with HAs observations; the alert rate, defined as the ratio of the number of alerts to the number of monitored visits; the intervention rate, defined as the proportion of alerts that led to health interventions; the response time, defined as the length of time from the day of the alert to the day of the intervention; and the nature of the health interventions. To analyze HAs adherence to the eHealth system application, we calculated the rate of HA-monitored visits as the ratio of the number of visits that resulted in observation records to the total number of HA visits recorded by the home care facility managers. Users' experience with this system was assessed according to Basch et al [21], using 2 self-administered anonymous questionnaires, one for the HAs and the facility managers, and one for the coordinating nurses who participated.

The secondary outcomes were to confirm the predictive capacity of the AI in real-world conditions. The diagnosis accuracy of the eHealth system alerts to predict ED visits was assessed by sensibility, specificity, positive and negative predictive values, and likelihood ratios. Accuracy analysis and reporting was conducted according to Standards for Reporting Diagnostics

Accuracy Studies guidelines. Occurrences and dates of ED visits and hospitalizations were recorded by the HAs in each visit and by the home aid facility manager.

### Statistical Analysis

Continuous variables were described by means and SD, or medians and IQRs, if not normally distributed; categorical variables were described by relative frequencies. The 2-tailed *t* test or Wilcoxon test was used to compare quantitative variables and the chi-square or Fisher exact test was used for qualitative variables. Sensitivity, specificity, positive and negative predictive value, and positive and negative likelihood ratios were estimated for each one of the alert visits (by an index test) in relation to an ED visit, that was considered the target condition of the reference standard. *P* values <.05 were considered statistically significant. Statistics were conducted using Stata software (version 16; StataCorp LLC).

### Ethics Approval

The research protocol was approved by the national French ethics committee for biomedical research, the Comité de Protection des Personnes, and the French Agency for the Safety of Health Products (2021-A02131-40–CPP 1-21-072 / 21.02093.000019).

Participants and the HAs and the home aid facilities' managers were informed about the nature and purpose of the study and provided their written consent accordingly.

## Results

### Participants and Home Aid Professionals Involved

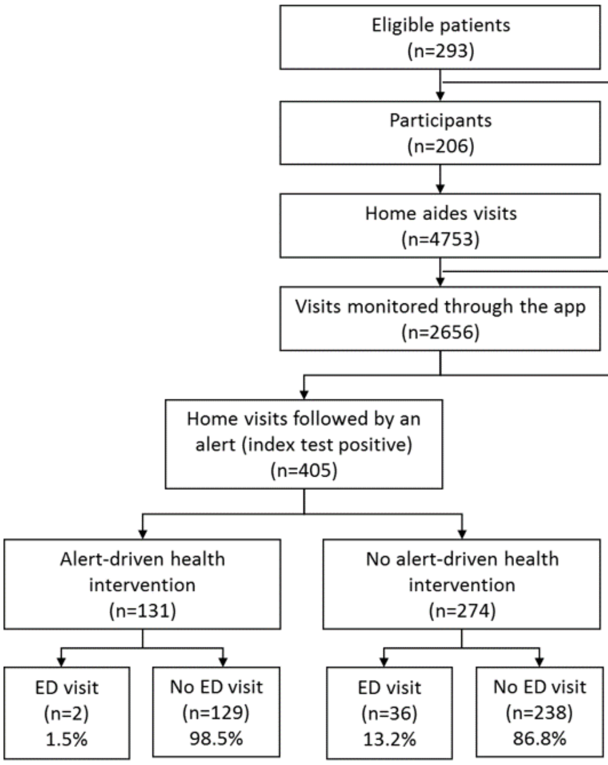
Among beneficiaries of the home aid facilities, 293 individuals were eligible, and 206/293 (71%) agreed to participate in the study and were included. Their mean age was 85 (SD 8) years, 161 of 206 (78%) participants were women, and 94 (45%) had a dependency level of GIR 3 or 4 (ie, moderate dependence level; Table 1).

The mean follow-up period was 10 months with no patient loss during the trial. In total, 10 care managers of the home aid facilities (9 nurses and 1 pharmacist) and 109 HAs were involved in the study. From the 4753 home visits, 2656 (56%) were monitored by the app and provided inputs for the eHealth system (Figure 2).

**Table 1.** Participants' characteristics, activity, and the eHealth system; staff involved in their functioning; and alert-triggered health interventions.

| Participants characteristics | Center 1 (n=67) | Center 2 (n=16) | Center 3 (n=123) | Total (N=206) |
|---|---|---|---|---|
| Age (years), mean (SD) | 86 (4) | 88 (6) | 86 (5) | 86 (5) |
| Gender (women), n (%) | 57 (86) | 13 (80) | 106 (86) | 176 (85) |
| Mild dependency (GIR 5 or 6), n (%) | 10 (15) | 3 (19) | 4 (3) | 17 (8) |
| Moderate dependency (GIR 3 or 4), n (%) | 23 (34) | 8 (50) | 63 (51) | 94 (46) |
| Severe dependency (GIR 1 or 2), n (%) | 0 (0) | 1 (6) | 13 (11) | 14 (7) |
| Unknown dependency level, n (%) | 34 (51) | 4 (25) | 43 (35) | 81 (39) |
| **eHealth system activity** | | | | |
| Home aides, n (%) | 46 (42) | 11 (10) | 52 (48) | 109 (100) |
| Care managers, n (%) | 6 (60) | 2 (20) | 2 (20) | 10 (100) |
| Visits monitored through the app, n (%) | 1130 (43) | 324 (12) | 1202 (45) | 2656 (100) |
| Compliance rate, % | 56.2 | 67·5 | 52.8 | 56.0 |
| Alerts, n (%) | 188 (46) | 47 (12) | 170 (42) | 405 (100) |
| Alert rate per monitored visits, % | 16.6 | 14.5 | 14.1 | 16.9 |
| **Alert-triggered health interventions** | | | | |
| Interventions, n (%) | 45 (34) | 46 (35) | 40 (31) | 131 (100) |
| Intervention rate per alerts, % | 23.9 | 97.9 | 23.5 | 32.3 |

**Figure 2.** The flowchart of alerts (index test), health interventions, and emergency department (ED) visits (reference standard).



## Emergency Visits During the Study

Of 206 participants, 29 (14%) visited EDs during the study. Of these, 11 made 2 or more visits (up to 6 visits), and the total number of ED visits was 46; a total of 32 ED visits (for 19 people) were followed by hospitalization.

Hospitalization by direct admission without passing through an ED visit was recorded in 5 participants (Figure 2).

## Health Interventions and ED Visits Occurring After an Alert Display

As a result of the 405 alerts generated by the eHealth system, 131 ATHI by health professionals were performed: 96 (73%) by nurses and 35 (27%) by physicians. After the 131 ATHI, only 2 ED visits (2/131, 1.5%) were recorded, whereas after the 272 alerts that did not result in a health intervention, we recorded 32 ED visits (13.2%), corresponding to an odds ratio of 0.10 (95% IC 0.02-0.43; *P*<.001; Table 2). These health interventions were performed by the patient's nurse or general practitioner.

Five patients died during the study. All had alerts, 4 did not have ATHI and were hospitalized, and 1 had an ATHI (*P*=.04; Table 3).

**Table 2.** Emergency department (ED) visits that occurred within 14 days of alerts generated by the eHealth system, according to the implementation of a health intervention triggered by the alerts.

| Characteristics | ED visits (n=38) | No ED visits (n=367) | Odds ratio (95% CI) |
|---|---|---|---|
| No alert-triggered health intervention, n (%) | 36 (13.1) | 238 (86.9) | Reference |
| Alert-triggered health intervention, n (%) | 2 (1.5) | 129 (98.5) | 0.10 (0.02-0.43)[a] |

[a] *P*<.001; *P*<.05 is considered statistically significant.

**Table 3.** Association between intervention and death.

| | Alert-triggered health interventions | | *P* value |
|---|---|---|---|
| | No (n=74) | Yes (n=132) | |
| **Death, n (%)** | | | .04 |
| No (n=201) | 70 (94.6) | 131 (99.2) | |
| Yes (n=5) | 4 (5.4) | 1 (0.8) | |

## Reports of Users' Experience

Users' experience surveys were completed by 81 of 109 (72%) HAs involved in the study and 8 of 10 (80%) coordinating nurses. In terms of understanding the approach, 83% (90/109) of HAs reported that the screening questions were easy to understand. In terms of overall usability, the digital system was easy to use for 90% (73/81) of HAs, and the response time was acceptable for 89% (72/81) of them (Figure 3).

The eHealth system was also well perceived by the coordinating nurses (Figure 4). Most of them found the app questions relevant; they believed the eHealth system had clinical utility and might improve interactions with patients and their family caregiver, and they mentioned that they would like to use it in the future and would recommend it to other facilities.

**Figure 3.** The satisfaction of home aides about the eHealth system.

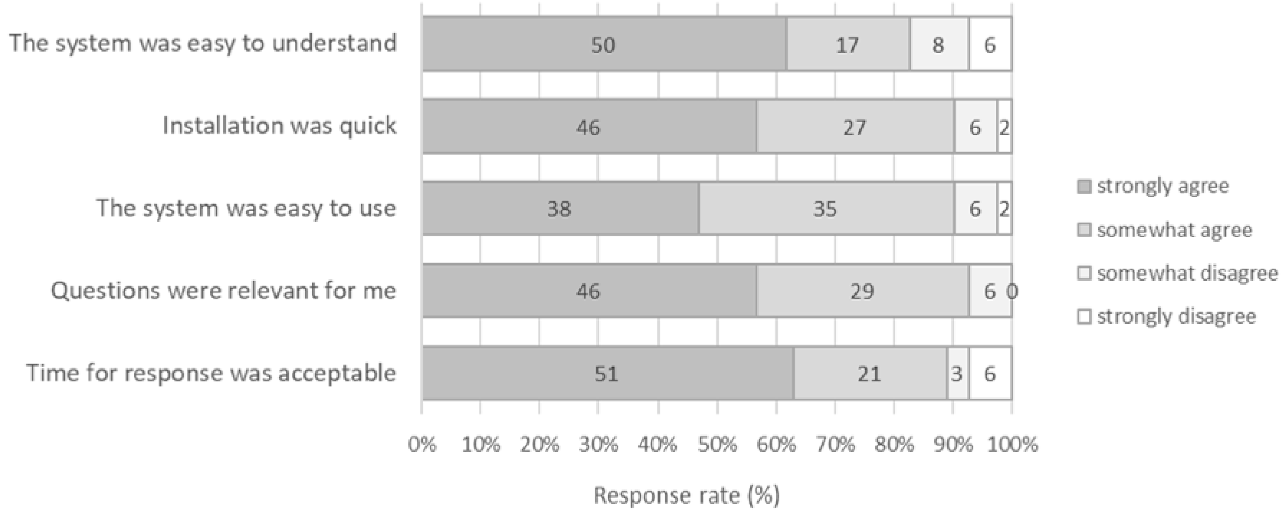
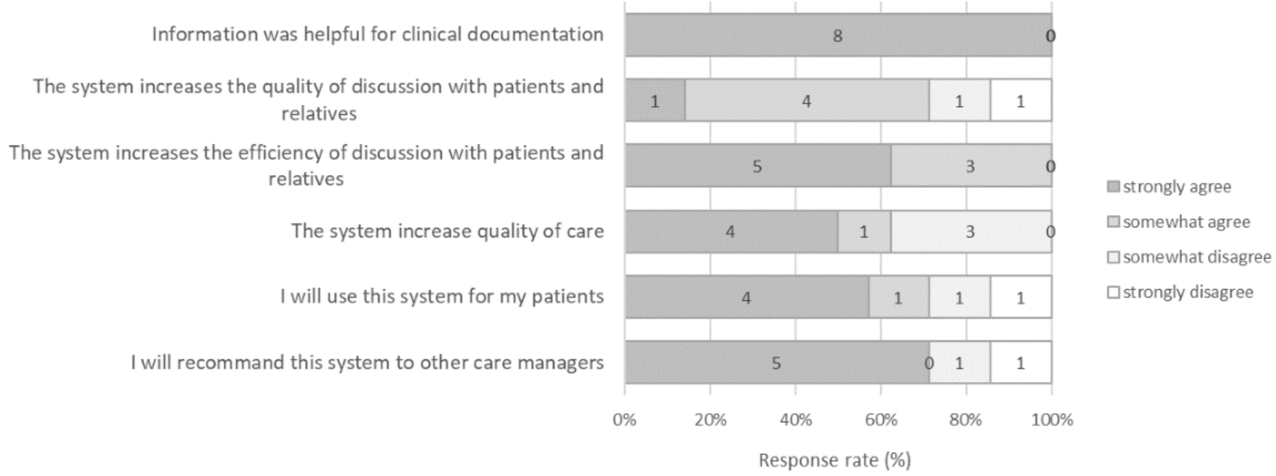


**Figure 4.** User's experience with 8 coordinating nurses who received the alerts and completed the questionnaire.



## The Alerts and Their Prediction of Emergency Department Visit

During the study, 405 alerts (between 22 and 49 per month) were displayed, corresponding to 15.2% (405/2656) of the monitored visits. Of the 46 ED visits, 8 (17%) were not preceded by alerts in the previous 14 days and 38 (82%) were preceded by alerts within that time frame (*P*<.001; Figure 1). The sensitivity and specificity of alerts for predicting ED visits that occurred within 14 days following the alerts were 83% (95% CI 72-94) and 86% (95% CI 85-87), respectively. The positive and negative predictive values were 9.4% (95% CI 6.5-12.2) and 99.6% (95% CI 99.3-99.9), respectively, and the positive and negative likelihood ratios were 5.87 (95% CI 4.99-6.92) and 0.20 (95% CI 0.11-0.38), respectively (Table 4 and 5).

**Table 4.** Contingency table for alerts generated by the eHealth system following home health aide visits and for emergency department visits occurring within 14 days of the alerts, and accuracy assessment.

| Characteristics | Home aides' visits with subsequent alerts (n=2656) | | *P* values |
|---|---|---|---|
| | Yes (n=405) | No (n=2251) | |
| Emergency department visits, n (%) | 38 (18.5) | 8 (0.3) | <.001 |
| No emergency department visits, n (%) | 367 (81.5) | 2243 (99.7) | <.001 |

**Table 5.** Characteristics of alerts for predicting emergency department visits.

| Home aides' visits with subsequent alerts (n=2656) | Accuracy assessment (95% CI) |
|---|---|
| Sensitivity, % (95% CI) | 83 (72-94) |
| Specificity, % (95% CI) | 86 (85-87) |
| Positive likelihood ratio | 5.87 (4.99-6.92) |
| Negative likelihood ratio | 0.20 (0.11-0.38) |
| Positive predictive value, % (95% CI) | 9.4 (6.5-12.2) |
| Negative predictive value, % (95% CI) | 99.6 (99.3-99.9) |

## Discussion

### Principal Results

In this intervention study, we successfully implemented an eHealth system based on HAs' observations and a prediction algorithm that is capable of informing health care professionals of the risk of an ED visit in the next two weeks. In total, 109 HAs were involved in the study for 4753 visits. More than half of the visits were monitored. Alerts automatically displayed by this eHealth system accurately predicted emergency room visits, and 32% (131/405) of them were followed by interventions by the patients' nurses or their general practitioners.

This eHealth system was well accepted and appreciated by HAs and their managers, and the accuracy was very good.

### Potential Bias and Levers

The evaluation of the accuracy of the alerts could have been biased by the transmission of the alert to a care manager. Nevertheless, it was shown that the interventions of the care managers allowed for a reduction in emergency hospitalizations, confirming the high predictive capacity of the system.

Balance between false negatives and false positives rates has been the subject of much reflection. In order to avoid a potential unnecessary ED visit (for a false positive, which is extremely rare), while avoiding as much as possible the loss of chances (false negative), the $F_1$-score has been chosen for the best optimization between false positives and false negatives rates.

### Comparison With Prior Studies

The eHealth system reported in this study overcame the classic obstacles faced by such systems. First, the completion of the smartphone-based, customer-centered diagnostic tool was good, and 90% (72/81) of the HAs found it acceptable. This is in contrast to studies that have highlighted that barriers to the use of e-PROMs for caregivers or clinicians are primarily related to long completion time and poor usability [22]. Second, acceptance of the alerts was satisfactory, and the health professionals who received them produced a high response rate for health interventions. This is probably related to the reasonable number of alerts and the ratio of alerts per visit that did not overwhelm practitioners, achieved through the good specificity of the machine learning algorithm. These results contrast with those of other studies that report that practitioners, overburdened by automatic alerts, no longer contact patients to intensify treatment of symptoms despite appropriate daily monitoring [21,23,24]. Other features of the machine learning algorithm have contributed to the acceptance of this eHealth system; its supervised nature provided health care professionals with indications of changes in beneficiaries' functional status (eg, ability to get up, move around, and eat, their mood, and loneliness) in addition to the alert alone; it helped them relieve their doubts and probably induced trust in the relevance of the alerts, since all the coordinating nurses found that information provided with alerts was clinically useful. It is likely that the acceptability rate would have been lower with deep learning algorithms that often have excellent predictive capabilities [25] because their operation is obscure to the users who receive the results, and that is a limitation when critical decisions need to be made.

Interestingly, in our study, the probability of an ED visit was very low after nurse or physician interventions following an alert, with a 10-fold decrease, compared to when alerts were not followed by such interventions. Even if this trial was not designed to examine such an outcome, this observation is very promising and prompts us to implement a controlled trial to document the effectiveness of the eHealth system. In addition, this device improves communication between professionals and promotes the empowerment of HAs. It responds to real public health issues for the prevention of the loss of autonomy in older people at home.

### Limitations

Our study faced some limitations. First, the study was conducted during the COVID-19 pandemic, which increased the risk of isolation for frail older adults and impacted primary care habits and HAs' working conditions. In addition, the incidence of ED visits was lower compared to previous studies of participants with similar characteristics (13% vs 40%). This may be due to factors other than our intervention, such as reticence to attend EDs for fear of exposure to COVID-19 [26]. This also raises the question of lack of data on hospitalization or death causes, which could have allowed a more detailed analysis according to the context (eg, COVID-19). Investigations are in progress to understand death causes and the patient's trajectory. Second, there is a limitation regarding deaths analysis; time of death was not taken into account, and it was not compared with the intervention's date. Therefore, a causal link between the two cannot be made. Third, for the moment, the alert is displayed only when it is reliable (ie, with enough data); however, to avoid a potential loss of chance, the person's condition and risks (geriatric and health) are transferred to the coordinating nurse, who can then assess the seriousness of the person's situation (ie, informing the coordinating nurse that it will be necessary to have other questionnaires to display an alert). Finally, this trial was not controlled, and a trial with a randomized controlled design should be conducted to document its clinical efficacy and cost-effectiveness.

### Perspectives

This study opens up broad prospects for optimizing the relevance of emergency visits for frail older adults. The predictive algorithm based on longitudinal observations of HAs could be improved by other types of input, such as patient's clinical or biological records or measurements from connected devices. In addition, our approach could be applied to new target events in older adults in specific health contexts, including oncogeriatrics, cardiogeriatrics, or postsurgery. Clinical investigations are currently in progress and will allow for investigating the transferability of this system to certain clinical contexts. Finally, the system might evolve to a decision support system to help health professionals to optimize and personalize ATHI.

### Conclusions

The eHealth system that we have successfully implemented offers an innovative approach to optimize the care of frail older adults. This approach is based on three paradigms: recording of the functional characteristics of daily life and their evolution over time, mobilization of nonprofessional health informants, and the use of a machine learning algorithm to monitor the level of individual risk and produce alerts that support health professional decisions for interventions. This means that multiple observers (not just social workers or nurses) could be trained to identify people at risk for ED visits. Such a predictive approach could form the basis for personalized health interventions that are designed to deliver early appropriate care and improve health outcomes.

## Acknowledgments

The authors would like to thank the managers and the staff of the following home aid facilities as well as the care managers for their contribution to the study: Adar Provence, Arcade Assistance, Association Le Connétable, Centre Communale d'Actions Sociales de la ville du Chesnay-Rocquencourt, Communauté Professionnels de Territoire de Santé Marseille Itinéraire Santé, Intervenant Libéraux et Hospitaliers Unis pour le Patient, Pôle Autonomie Territorial Grand Versailles, and Plateforme Territoriale d'Appui Marseillaise.

This work is funded by the Agence Régionale de Santé Provence Alpes Cotes d'Azur and the Agence du Numérique en Santé. They did not play any role in the study except for the participation of Agence Régionale de Santé Provence Alpes Cotes d'Azur in the discussion about the ATHI.

## Data Availability

The data that support the findings of this study are available from the corresponding author, JHV, upon reasonable request. Restrictions apply to the availability of the data collected by the home aids and the algorithm that was used in this study, which is the subject of a patent.

## Authors' Contributions

Each author met the International Committee of Medical Journal Editors criteria for authorship. JB, JHV, and FD conceptualized the study; JB, PV, MG, and SF conducted the acquisition of data; JB, JHV, FD, and PV contributed to the interpretation of the results; JB; JHV, and CHT were in charge of the formal analysis; JB, JHV, CHT, SM, FD, and PV drafted and revised the manuscript. All authors approved the final version of the manuscript.

## Conflicts of Interest

JB received funding from Pfizer and Novartis for conferences and board participation. FD reports conflicts of interests with Chugai, Astra-Zeneca, Merck, Sivan, Takeda, Ipsen, Bristol Meyer Squibb, Viatris, Kelindi, and Hyperion where he is an invited speaker, and also serves on the advisory board at Sivan and Roche. The author holds stocks and shares with INeS, Kelindi, Hyperion and the institution receiver was Hyperion. JHV is the director of the company PRESAGE. CHT is employed by Présage care as a research and innovation coordinator. Other authors do not report any conflicts of interest.

Multimedia Appendix 1
The list of 23 items recorded by home care aides at each home visit and their completeness rates.
[DOCX File , 18 KB - jmir_v24i9e40387_app1.docx ]

Multimedia Appendix 2
A simplified diagram of the processing system and the description of data processing system.
[DOCX File , 53 KB - jmir_v24i9e40387_app2.docx ]

## References

1. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet 2012 Jul;380(9836):37-43. [doi: 10.1016/S0140-6736(12)60240-2]
2. Prince MJ, Wu F, Guo Y, Gutierrez Robledo LM, O'Donnell M, Sullivan R, et al. The burden of disease in older people and implications for health policy and practice. The Lancet 2015 Feb;385(9967):549-562. [doi: 10.1016/S0140-6736(14)61347-7]
3. Hwang U, Morrison RS. The geriatric emergency department. J Am Geriatr Soc 2007 Nov;55(11):1873-1876. [doi: 10.1111/j.1532-5415.2007.01400.x] [Medline: 17916122]
4. Duong HV, Herrera LN, Moore JX, Donnelly J, Jacobson KE, Carlson JN, et al. National characteristics of emergency medical services responses for older adults in the United States. Prehosp Emerg Care 2018 Sep 01;22(1):7-14 [FREE Full text] [doi: 10.1080/10903127.2017.1347223] [Medline: 28862480]
5. Chen C, Wang C, Huang G. Functional trajectory 6 months posthospitalization: a cohort study of older hospitalized patients in Taiwan. Nurs Res 2008;57(2). [doi: 10.1097/01.nnr.0000313485.18670.e2]
6. Iwashyna TJ, Ely EW, Smith DM, Langa KM. Long-term cognitive impairment and functional disability among survivors of severe sepsis. JAMA 2010 Oct 27;304(16):1787-1794 [FREE Full text] [doi: 10.1001/jama.2010.1553] [Medline: 20978258]
7. Adams JG. Emergency department overuse: perceptions and solutions. JAMA 2013 Mar 20;309(11):1173-1174. [doi: 10.1001/jama.2013.2476] [Medline: 23512065]
8. McCusker J, Verdon J. Do geriatric interventions reduce emergency department visits? a systematic review. J Gerontol A Biol Sci Med Sci 2006 Jan;61(1):53-62. [doi: 10.1093/gerona/61.1.53] [Medline: 16456194]
9. Gasperini B, Cherubini A, Pierri F, Barbadoro P, Fedecostante M, Prospero E. Potentially preventable visits to the emergency department in older adults: results from a national survey in Italy. PLoS One 2017 Dec 21;12(12):e0189925 [FREE Full text] [doi: 10.1371/journal.pone.0189925] [Medline: 29267333]
10. Uscher-Pines L, Pines J, Kellermann A, Gillen E, Mehrotra A. Emergency department visits for nonurgent conditions: systematic literature review. Am J Manag Care 2013 Jan;19(1):47-59 [FREE Full text] [Medline: 23379744]
11. Fingar KR, Barrett ML, Elixhauser A, Stocks C, Steiner CA. Trends in potentially preventable inpatient hospital admissions and emergency department visits. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs [Internet]. 2015. URL: https://www.ncbi.nlm.nih.gov/books/NBK338139/ [accessed 2018-12-25]
12. Bouazza YB, Chiairi I, El Kharbouchi O, De Backer L, Vanhoutte G, Janssens A, et al. Patient-reported outcome measures (PROMs) in the management of lung cancer: a systematic review. Lung Cancer 2017 Nov;113:140-151. [doi: 10.1016/j.lungcan.2017.09.011] [Medline: 29110842]
13. Denis F, Krakowski I. How should oncologists choose an electronic patient-reported outcome system for remote monitoring of patients with cancer? J Med Internet Res 2021 Sep 09;23(9):e30549 [FREE Full text] [doi: 10.2196/30549] [Medline: 34499046]
14. Schick-Makaroff K, Karimi-Dehkordi M, Cuthbertson L, Dixon D, Cohen SR, Hilliard N, et al. Using patient- and family-reported outcome and experience measures across transitions of care for frail older adults living at home: a meta-narrative synthesis. Gerontologist 2021 Apr 03;61(3):e23-e38 [FREE Full text] [doi: 10.1093/geront/gnz162] [Medline: 31942997]
15. Veyron J, Friocourt P, Jeanjean O, Luquel L, Bonifas N, Denis F, et al. Home care aides' observations and machine learning algorithms for the prediction of visits to emergency departments by older community-dwelling individuals receiving home care assistance: a proof of concept study. PLoS One 2019;14(8):e0220002 [FREE Full text] [doi: 10.1371/journal.pone.0220002] [Medline: 31408458]
16. Huntley AL, Chalder M, Shaw ARG, Hollingworth W, Metcalfe C, Benger JR, et al. A systematic review to identify and assess the effectiveness of alternatives for people over the age of 65 who are at risk of potentially avoidable hospital admission. BMJ Open 2017 Aug 01;7(7):e016236 [FREE Full text] [doi: 10.1136/bmjopen-2017-016236] [Medline: 28765132]
17. Soril LJJ, Leggett LE, Lorenzetti DL, Noseworthy TW, Clement FM. Reducing frequent visits to the emergency department: a systematic review of interventions. PLoS One 2015 Apr 13;10(4):e0123660 [FREE Full text] [doi: 10.1371/journal.pone.0123660] [Medline: 25874866]

18. Nord M, Lyth J, Alwin J, Marcusson J. Costs and effects of comprehensive geriatric assessment in primary care for older adults with high risk for hospitalisation. BMC Geriatr 2021 Apr 21;21(1):263 [FREE Full text] [doi: 10.1186/s12877-021-02166-1] [Medline: 33882862]
19. Belmin J, Auffray J, Berbezier C, Boirin P, Mercier S, de Reviers B, et al. Level of dependency: a simple marker associated with mortality during the 2003 heatwave among French dependent elderly people living in the community or in institutions. Age Ageing 2007 May 24;36(3):298-303. [doi: 10.1093/ageing/afm026] [Medline: 17384419]
20. Haute ADS. Cartographie des impacts organisationnels pour l'évaluation des technologies de santé. Haute Autorité de santé. 2020 Dec 31. URL: https://www.has-sante.fr/jcms/c_2902770/fr/cartographie-des-impacts-organisationnels-pour-l-evaluation-des-technologies-de-sante [accessed 2022-02-27]
21. Basch E, Stover AM, Schrag D, Chung A, Jansen J, Henson S, et al. Clinical utility and user perceptions of a digital system for electronic patient-reported symptom monitoring during routine cancer care: findings from the PRO-TECT trial. JCO Clin Cancer Inform 2020 Oct;4:947-957 [FREE Full text] [doi: 10.1200/CCI.20.00081] [Medline: 33112661]
22. Duncanson E, Bennett PN, Viecelli A, Dansie K, Handke W, Tong A, Symptom monitoring WIth Feedback Trial (SWIFT) Investigators. Feasibility and acceptability of e-PROMs data capture and feedback among patients receiving haemodialysis in the Symptom monitoring WIth Feedback Trial (SWIFT) pilot: protocol for a qualitative study in Australia. BMJ Open 2020 Nov 06;10(11):e039014 [FREE Full text] [doi: 10.1136/bmjopen-2020-039014] [Medline: 33158824]
23. Adam R, Burton CD, Bond CM, de Bruin M, Murchie P. Can patient-reported measurements of pain be used to improve cancer pain management? A systematic review and meta-analysis. BMJ Support Palliat Care 2017 Dec 22;7(4). [doi: 10.1136/bmjspcare-2016-001137] [Medline: 27879472]
24. Anderson KO, Palos GR, Mendoza TR, Cleeland CS, Liao K, Fisch MJ, et al. Automated pain intervention for underserved minority women with breast cancer. Cancer 2015 Jun 01;121(11):1882-1890 [FREE Full text] [doi: 10.1002/cncr.29204] [Medline: 25711974]
25. Beaudouin V, Bloch I, Bounie D, Clémençon S, d'Alché-Buc F, Eagan J, et al. Flexible and context-specific AI explainability: a multidisciplinary approach. SSRN Journal 2020 Apr 10:1-66. [doi: 10.2139/ssrn.3559477]
26. Howley F, Lavan A, Connolly E, McMahon G, Mehmood M, Briggs R. Trends in emergency department use by older people during the COVID-19 pandemic. Eur Geriatr Med 2021 Dec 17;12(6):1159-1167 [FREE Full text] [doi: 10.1007/s41999-021-00536-x] [Medline: 34273092]

## Abbreviations

**ATHI:** alert-triggered health intervention
**ED:** emergency department
**HA:** home aide
**PROM:** patient-reported outcome measure

*Edited by G Eysenbach; submitted 18.06.22; peer-reviewed by W Klement, B Li, H Musawir; comments to author 21.07.22; revised version received 28.07.22; accepted 30.07.22; published 08.09.22.*

*Please cite as:*
*Belmin J, Villani P, Gay M, Fabries S, Havreng-Théry C, Malvoisin S, Denis F, Veyron JH*
*Real-world Implementation of an eHealth System Based on Artificial Intelligence Designed to Predict and Reduce Emergency Department Visits by Older Adults: Pragmatic Trial*
*J Med Internet Res 2022;24(9):e40387*
*URL: https://www.jmir.org/2022/9/e40387*
*doi:10.2196/40387*
*PMID:35921685*

Original Paper

# The Data-Adaptive Fellegi-Sunter Model for Probabilistic Record Linkage: Algorithm Development and Validation for Incorporating Missing Data and Field Selection

Xiaochun Li[1*], PhD; Huiping Xu[1*], PhD; Shaun Grannis[2], MD

[1]Department of Biostatistics and Health Data Science, Indiana University School of Medicine, The Richard M. Fairbanks School of Public Health, Indianapolis, IN, United States

[2]Data and Analytics, Regenstrief Institute Inc., Indiana University School of Medicine, Indianapolis, IN, United States

[*]these authors contributed equally

**Corresponding Author:**
Xiaochun Li, PhD
Department of Biostatistics and Health Data Science
Indiana University School of Medicine
The Richard M. Fairbanks School of Public Health
HITS, Suite 3000
410 W 10th St.
Indianapolis, IN, 46202
United States
Phone: 1 317 274 2696
Email: xiaochun@iu.edu

## Abstract

**Background:** Quality patient care requires comprehensive health care data from a broad set of sources. However, missing data in medical records and matching field selection are 2 real-world challenges in patient-record linkage.

**Objective:** In this study, we aimed to evaluate the extent to which incorporating the missing at random (MAR)–assumption in the Fellegi-Sunter model and using data-driven selected fields improve patient-matching accuracy using real-world use cases.

**Methods:** We adapted the Fellegi-Sunter model to accommodate missing data using the MAR assumption and compared the adaptation to the common strategy of treating missing values as disagreement with matching fields specified by experts or selected by data-driven methods. We used 4 use cases, each containing a random sample of record pairs with match statuses ascertained by manual reviews. Use cases included health information exchange (HIE) record deduplication, linkage of public health registry records to HIE, linkage of Social Security Death Master File records to HIE, and deduplication of newborn screening records, which represent real-world clinical and public health scenarios. Matching performance was evaluated using the sensitivity, specificity, positive predictive value, negative predictive value, and F1-score.

**Results:** Incorporating the MAR assumption in the Fellegi-Sunter model maintained or improved F1-scores, regardless of whether matching fields were expert-specified or selected by data-driven methods. Combining the MAR assumption and data-driven fields optimized the F1-scores in the 4 use cases.

**Conclusions:** MAR is a reasonable assumption in real-world record linkage applications: it maintains or improves F1-scores regardless of whether matching fields are expert-specified or data-driven. Data-driven selection of fields coupled with MAR achieves the best overall performance, which can be especially useful in privacy-preserving record linkage.

***(J Med Internet Res 2022;24(9):e33775)*** doi:10.2196/33775

**KEYWORDS**

record linkage; Fellegi-Sunter model; latent class model; missing at random; matching field selection

## Introduction

Quality patient care requires comprehensive health care data from a broad set of sources. Electronic medical record (EMR) data are increasingly distributed across many sources as the era of digital health care is accelerated in the United States. However, EMR data from independent databases often lack a common patient identifier, which impedes data aggregation, causes inefficiencies (eg, tests repeated unnecessarily), affects patient care, and hinders research. Record linkage is a requisite step for effective and efficient patient care and research. Without a unique universal patient identifier, linkage of patient records is a nontrivial task. The simplest class of approaches is the deterministic method, which requires the strict identity of the selected data elements of a pair of records, such as name, birthdate, gender, and Social Security number. Although deterministic algorithms are generally simple to implement and achieve excellent specificity, they have low sensitivity, are not robust to missing data, cannot quantify the uncertainty of the matching process, and are inflexible to changing data characteristics.

The Fellegi-Sunter (FS) [1] model is widely used for probabilistic record linkage based on the binary agreement or disagreement of a select set of fields of record pairs, such as Social Security number, first name, middle name, last name, and date of birth. The FS model is in essence a latent class model applied to record linkage problems. The latent class variable is the unobserved true match status, and the parameters in the model are the match prevalence, probabilities of field agreements among true matches (m-probabilities), and probabilities of field agreements among nonmatches (u-probabilities). A record pair's matching weights are defined as the logarithms of the m- and u-probability ratios, and the sum of the weights is the matching score of the pair. Record pairs were then classified into matches and nonmatches based on their matching scores for a given threshold. The linking algorithm based on the FS model is shown to outperform the deterministic algorithm [2]. However, methodological gaps exist in configuring and applying the FS model.

First, it is well known that missing data are prevalent in real-world data in EMRs [3]. Data necessary for matching records are often missing from clinical data for many reasons: values may be coded as "unknown," nonexistent (a person with no middle name), or omitted due to privacy concerns (such as Social Security number). Missing field values decrease the information content in the data and consequently hinder matching accuracy. Matching only records with full information is undesirable because it excludes many records and thus misses matches. One study found that mother's date of birth was often absent because it was not the focus of pediatricians' attention [4]. However, this information significantly improved the linkage procedure when present. Therefore, effective accommodation of missing data is needed to maximize linkage. Common strategies in practice involve excluding records with missing values in any of the matching fields when estimating match weights [5] or considering the missing field's agreement pattern as disagreement [6] (missing as disagreement [MAD]). The former lacks efficiency because of the loss in sample size due to exclusion. The latter does not account for the fact that true matches can contain missing fields and is deficient in a theoretical justification. Another strategy is to model missing data in a matching field as the third category, in addition to the categories of agree and disagree [7]. However, it is well established that including missing data by adding a category "missing" causes serious biases, even when data are missing completely at random [8-13]. In a model-based approach, Enamordo et al [14] assume that data in matching fields are missing at random (MAR) conditional on the true match status. Their comprehensive simulation studies show that the FS model with MAR incorporated outperforms deterministic linkage in social science when linking voter files. How the FS model with MAR incorporated compares with the FS model using zero-filled data in which missing values in the original data are replaced by 0 by MAD has not been evaluated. Furthermore, while MAR is evaluated and applied to voting registries, its performance in linking EMR files is not known.

Second, although there may be numerous fields (or attributes) across record files not all of them are useful for matching. For example, if matching 2 obstetrics and gynecology databases, the field "gender" is not informative. In real-world data, there are likely also dependencies among the data fields. As we have demonstrated [15], the FS model exhibits poor matching accuracy when the fields are highly correlated. As more fields are used in the FS model, more dependencies may be introduced. Ideally, the FS model should be able to use a minimally sufficient set of fields. However, we are unaware of data-driven methods for matching field selection. In practice, the expert input is solicited to identify an appropriate subset matching fields. Several iterations may be required to achieve the desired match accuracy using a manually reviewed data set with known match statuses among record pairs. This process is neither scalable nor generalizable and is infeasible in privacy-preserving record linkage [7]. We are also unaware of any work that evaluates the effects of missing data treatment and field selection for matching simultaneously.

We will evaluate the effects of incorporating missing data treatment and matching field selection into the FS algorithm on linkage performance using 4 real-world use cases in our local operational data aggregation system—a health information exchange (HIE) environment, into which different data sources are integrated. The 4 use cases included health information exchange record deduplication (labeled as Indiana Network for Patient Care [INPC]), linkage of a public health registry Marion County Health Department records to HIE (labeled as MCHD), linkage of Social Security Death Master File records of the Social Security Administration to HIE (labeled as SSA), and deduplication of newborn screening records (labeled as NBS). We hypothesize that proper treatment of missing data and data-driven matching field selection will enhance linkage performance.

## Methods

### Blocking

Records need to be compared in record linkage to ascertain whether they belong to the same entity. Forming record pairs by Cartesian product from the 2 files (or to a file itself in the case of deduplication) results in an enormously large number of pairs. For example, the data set from the INPC (the INPC use case) has 47,334,986 records (Table 1) and will form 2.24 quadrillion record pairs by the Cartesian product. A common strategy is “blocking on” certain fields (blocking variables) to reduce the number of record pairs; that is, retaining only those record pairs with exact agreement in blocking variables. Blocking helps to enrich matches by restricting the search space. We applied 5 blocking schemes to each use case. In the INPC use case, the five blocking schemes are the Social Security number (SSN); first name and telephone number (FN-TEL); day, month, and year of birth and zip code (DB-MB-YB-ZIP); first name, last name, and year of birth (FN-LN-YB); and day, month, and year of birth and last name (DB-LN-MB-YB). These five blocking schemes contained 613 million record pairs, with the number of pairs in each block listed in Table 1. Within each block, record pairs are compared field by field for a collection of matching fields, yielding a vector of comparison results for each pair. For example, if only 3 matching fields are compared by exact comparison (for agreement or disagreement), the vectors will have $2^3$ possible patterns when there are no missing data. In general, if $K$ matching fields are compared, there will be $2^K$ total agreement patterns.

**Table 1.** Summary of four use cases, Indiana Network for Patient Care (INPC), newborn screening (NBS), Social Security Administration (SSA), and Marion County Health Department (MCHD), with information on the number of records in each use case, blocking schemes, and the numbers of record pairs in blocking schemes.

| Block | Pairs |
|---|---|
| **INPC (47,334,986 records)** | |
| SSN[a] | 53,054,690 |
| FN-TEL[b] | 41,729,402 |
| DB-MB-YB-ZIP[c] | 133,553,036 |
| FN-LN-YB[d] | 193,865,283 |
| DB-LN-MB-YB[e] | 191,181,498 |
| **NBS (765,813 records)** | |
| MRN[f] | 4,147,098 |
| TEL[g] | 2,644,454 |
| MB-DB-ZIP | 8,083,396 |
| LN-FN[h] | 3,005,368 |
| NK_LN-NK_FN[i] | 1,217,736 |
| **SSA (89,556,520 records)** | |
| SSN | 805,331 |
| FN-LN-ZIP | 18,103 |
| FN-LN-MI-YB | 1,395,395 |
| FN-LN-MI-DB-MB | 547,376 |
| FN-LN-DB-MB-YB | 722,167 |
| **MCHD (471,298 records)** | |
| SSN | 869,454 |
| TEL | 28,238 |
| DB-MB-YB-zip | 5,083,429 |
| FN-LN-YB | 3,378,017 |
| DB-LN-MB-YB | 3,701,460 |

[a] SSN: Social Security number.

[b] FN-TEL: first name and telephone number.

[c] DB-MB-YB-ZIP: day, month, and year of birth and zip code.

[d] FN-LN-YB: first name, last name, and year of birth.

[e] DB-LN-MB-YB: day, month, and year of birth and last name.

[f] MRN: medical record number.

[g] TEL: telephone no.

[h] LN-FN: last name, first name.

[i] NK_LN, NK_FN: next of kin last name and first name.

## The FS Model

Formally, for the *i*th pair of records, let $\delta_i$ denote the unobserved true match status (a latent binary class variable) with a value of 1 indicating a match and 0 indicating a nonmatch (ie, the class label for match and nonmatch classes), $Y_i=(Y_{i1}, \dots, Y_{iK})$ be the vector of agreements in K fields, and $y_i=(y_{i1}, \dots, y_{iK})$ be the observed agreements. In addition, let *n* be the total number of pairs and $\rho=P(\delta=1)$, the *match prevalence* in the total n pairs of records. Assuming independent observations $(y_i, \delta_i)$, $i=1,...,n$, we express its complete data likelihood and the marginal distribution of $y_i$, $i=1,...,n$ as follows:

and . For a given *i*, the posterior probability of $\delta_i=1$ is . If the true match status $\delta_i$'s are known, then the MLE of $\rho$ for the complete data likelihood is . When $\delta_i$'s are unknown, this problem is known as the latent class modeling because the model parameters are estimated without the class label being observed.

A popular algorithm, named after Fellegi and Sunter [1] in the probabilistic record linkage literature, further assumes ; that is, the assumption of conditional independence of $Y_{i1}$, ... ,$Y_{iK}$ within each latent class. The FS model greatly simplifies the estimation process, producing estimates for field-specific probability of agreement given that a pair is a match, $m_k=P(Y_{ik}=1|\delta_i=1)$, and the field-specific probability of agreement given that a pair is a nonmatch, $u_k=P(Y_{ik}=1|\delta_i=0)$. Model estimates can be obtained by using the Expectation-Maximization (EM) algorithm [16] on the complete data likelihood or by using standard optimization routines on the marginal likelihood. The FS approach allows the model parameters to be estimated based on the observed agreements of pairs *without* the use of a training set, qualifying it as an unsupervised learning algorithm.

## Classification of Record Pairs

Match scores are defined as the logarithm of likelihood ratios, . Under the conditional independence assumption, the match score for the *i*th pair is the sum of the logarithms of the field-specific likelihood ratios, . Match scores are computed using the estimated $m_k$ and $u_k$ from the FS model and are in turn used to rank all record pairs, with a high score indicating a higher likelihood of a record pair to be a match. In our study, we used the estimated match prevalence ρ to set the threshold as the upper ρ-th quantile of the scores. A record pair is then declared as a match if its score is greater than the threshold; otherwise, it is declared as a nonmatch.

## Treatment of Missing Data

Formally describing the missing data mechanism is important for devising an approach to account for missing data. Missing data are generally classified into 3 types [17]. First, the most restrictive type of missing data is missing completely at random (MCAR), which assumes that the missingness in a variable is independent of all observed or unobserved variables. In this situation, the parameter estimates are unbiased when record pairs with any missing data are excluded. However, omitting missing data may lower the precision of estimated parameters due to the smaller sample size. In addition, MCAR is a strong assumption that cannot be verified with the data at hand. Second, MAR is a less-restrictive yet more realistic missing data model that assumes that the missingness in a variable is independent of unobserved data, although it can depend on other observed variables. Finally, missing not at random (MNAR) asserts that the missingness of a variable is related to the unobserved variable itself. To handle MNAR, knowledge of the missing mechanism is required to model the missing process in the estimation of the parameters and matching scores.

In record linkage applications, missing values in matching fields are typically handled by excluding records with missing values on one of the matching fields when estimating match weights [5] or considering the field's agreement pattern as a disagreement [6]. Excluding records with missing values is justifiable only when the data are MCAR. Thus, excluding records when the MCAR assumption does not hold leads to inaccurate results due to bias and low precision; the bias arises from the wrong model assumption and the low precision from the reduced sample size. Alternatively, treating missing data as disagreement (MAD) is implicitly invoking the assumption of MNAR, which may yield inaccurate results when the MAD assumption that all missing data represent disagreement is incorrect. This strong assumption is likely false for data to be linked. For example, if the middle name is absent because it does not exist, a missing value from both records of the record pair can provide information that the 2 records belong to the same person. On the other hand, the assumption of MAR is the least restrictive among the 3 types of missing mechanisms, and we hypothesize that it will yield superior match performance. Assuming MAR, the missing data are handled using the full information likelihood approach that uses all available data (ignoring the matching fields with missing values) in the FS model under the assumption of conditional independence of the matching fields.

The predictive results are obtained the same way for the FS model with MAR and MAD. The difference lies in the manner in which missing data are treated. When MAD is used, fields with missing data are set to "disagreement" (coded as 0), and the FS algorithm as is can proceed on the data with missing values replaced by zeros. When MAR is used, the FS algorithm is used on nonmissing data. In either cases, parameters $m_k$,$u_k$ and the match prevalence are estimated, and match scores are calculated for all pairs. The threshold for a pair to be a match is set to be the upper -th quantile of the scores. A record pair is then declared as a match if its score is greater than the threshold; otherwise, it is declared a nonmatch.

## Selection of Matching Fields

Fields missing 100% within a blocking scheme contain no information and will not be considered further. We examined 2 approaches selecting matching fields: the standard practice of subject matter expert-guided field selection and a data-driven approach. In the data-driven approach, all fields were considered to be putative matching fields. A necessary condition for a field to be useful in matching is that it should exhibit variability. For example, if the value of a field is fixed (no variation), it cannot separate matches from nonmatches. Thus, a blocking variable can no longer be used as a matching field in a block formed using the blocking variable. When running an FS model, we started with the largest possible set of fields; more fields may be dropped from the model, starting with fields with the least variations, until the FS algorithm converges.

## Data Sets of 4 Use Cases and Gold Standards

We evaluated the matching performance of the missing data treatment (MAD and MAR) and matching field selection (expert-specified fields vs data-driven fields) by conducting a 2-by-2 factorial design using 4 real-world use cases in our local HIE environment. The 4 use cases contain data that were generated as part of clinical or public health processes.

The 4 use cases included deduplicating clinical records in a state-level HIE, linking a public health registration file to clinical data in the HIE, linking death records to clinical data in the HIE, and deduplication of the Health Level Seven International (HL7) messages for newborns less than 1 month of age from the HIE. For each use case, blocking was performed to confine the total number of record pairs to be compared with a subspace of record pairs enriched with true matches [18]. Five blocking schemes were selected for each use case based on expert input from our laboratory. The total number of records for each use case and the number of record pairs per blocking scheme are listed in Table 1. To assess the matching performance, we selected record pairs for human review by performing proportional sampling from the union of record pairs with strata defined by the five blocking schemes. To compare the sensitivity of the algorithms, a total of 5884 true matches were necessary to test a 2% absolute difference in discordant rates of the 2 algorithms among the true matches with an 80% power at a 2-sided significance level of .05. The same number of true nonmatches was required to test the 2% difference in specificity. We sampled record pairs until we reached at least 5884 pairs in the class of true matches and the class of true nonmatches. Each record pair was reviewed by 2 reviewers; in the case of a disagreement in the classification of the pair, a third reviewer adjudicated the pair. Table 2 summarizes the manually reviewed sets for the 4 use cases.

**Table 2.** Manual review results for the 4 use cases.

| Use case | Number of pairs[a] | Number of pairs deemed as matches | Number of pairs deemed as nonmatches | Match prevalence[b] |
|---|---|---|---|---|
| INPC[c] | 15,000 | 7840 | 7160 | 0.523 |
| SSA[d] | 16,500 | 5950 | 10,550 | 0.361 |
| NBS[e] | 15,000 | 7967 | 7033 | 0.531 |
| MCHD[f] | 15,500 | 5927 | 9573 | 0.382 |

[a] Number of pairs is the total number of pairs sampled for manual review, which determines the pairs as either matches or nonmatches.

[b] Match prevalence is the ratio of the number of pairs deemed as matches and the total number of pairs for manual review for each use case.

[c] INPC: Indiana Network for Patient Care.

[d] SSA: Social Security Administration.

[e] NBS: newborn screening.

[f] MCHD: Marion County Health Department.

### *Deduplicating HIE (INPC)*

This data set reflected demographic records from geographically proximal hospital systems that participate in HIE. Blocking is as described earlier. The data contained a subset of 15,000 sampled gold standard pairs with 7840 (52.3%) true positives and 7160 (47.7%) true negatives. Patients from hospitals in close proximity cross over to nearby institutions, creating the need to identify common records. New value-based purchasing models such as Accountable Care Organizations dramatically increased the need to identify and capture information on patients seeking care from other institutions.

### *HIE and Vital Records for Ascertaining Death Status (SSA)*

These data reflect a combination of the Social Security Death Master File and HIE data. We applied five blocking schemes (Table 1): SSN; first name, last name, and zip code (FN-LN-ZIP); first name, last name, middle initial, and year of birth (FN-LN-MI-YB); first name, last name, middle initial, and day and month of birth (FN-LN-MI-DB-MB); and first name, last name, and day, month, and year of birth (FN-LN-DB-MB-YB). This data set contained a subset of 16,500 sampled gold standard pairs with 5950 (36.1%) true positives and 10,550 (63.9%) true negatives. Accurately and comprehensively updating health records with patients' accurate death status is critical for robust clinical quality measurement, public health reporting requirements, and high-quality clinical research.

### *Deduplicating Newborn Registration Data (NBS)*

This data set included demographic data for newborns derived from multiple hospitals, clinics, and within the HIE. These data were limited to patients aged <2 months. We applied five blocking schemes (Table 1): medical record number (MRN), telephone number (TEL), month, day of birth, zip code (MB-DB-ZIP), last name and first name (LN-FN), and next of kin's last name and first name (NK_LN-NK_FN). This data set contained a subset of 15,000 sampled gold standard pairs, with 7967 (53.1%) true positives and 7033 (46.9%) true negatives. Matching in this cohort is important because not all infants receive appropriate screening for harmful or potentially fatal disorders that are otherwise unapparent at birth [4]. Public health screening tests must be linked to patient records to avoid harmful delays in diagnosis.

### *Public Health Registry Linked to Clinical Registrations (MCHD)*

This data set comes from the MCHD, Indiana's largest public health department. The registry contains a master list of demographic information for clients who receive public health services such as immunization; Women, Infants, and Children's nutrition support; and laboratory testing [19,20]. The registry also tracks population health trends and supports other public health activities. Duplicate patient records are often unintentionally added. We applied five blocking schemes (Table 1): SSN; telephone number (TEL); day, month, and year of birth and zip code (DB-MB-YB-ZIP); first name, last name, and year of birth (FN-LN-YB); and day, month, and year of birth and last name (DB-LN-MB-YB). This data set contained a subset of 15,500 sampled gold standard pairs with 5927 (38.2%) true positives and 9573 (61.8%) true negatives. We linked the complete patient registry to patient records in the aforementioned HIE.

The 4 data sets contained subsets of the following fields: MRN, SSN, last name (LN), first name (FN), middle initial (MI), nickname (NICK_SET), ethnicity (ETH_IMP), sex, month of birth (MB), day of birth (DB), YB, street address (ADR), city, state (ST), zip code (ZIP), telephone number (TEL), email, last name of next of kin (NK_LN), first name of next of kin (NK_FN), last name of treating physician (DR_FN), and first name of treating physician (DR_LN). The last 4 fields were used only in the NBS use case.

### Analyses of Use Cases

For each use case, blocking was performed first, and five blocks of record pairs were generated. The blocking schemes are listed in Table 1. The FS model is applied 4 times in each block based on the 2-by-2 factorial design, where missing data are either treated using MAD or MAR, and matching fields are either expert-specified or selected by the data-driven method. The parameters of the FS model can be estimated using the Newton Raphson approach or the EM algorithm, both of which maximize the likelihood function of the model. The exact agreement on the following fields (when available for a use case) was considered in the matching in INPC, MCHD, and SSA use cases: street address, city (in address), DB, MB, YB, EMAIL, ethnicity, FN, LN, MI, MRN, nickname, sex, state, ZIP, and TEL. MCHD and SSA do not have MRN; all 4 use cases include the nickname and ethnicity as derived fields. The fields used in the NBS use case were slightly different. The exact agreement on the following fields was considered in the matching in the NBS use case: street address, city, DB, MB, YB, physician's FN, physician's LN, email address, ethnicity, FN, LN, MI, nick name, MRN, NK_FN, NK_LN, sex, SSN, state, TEL, and ZIP. The fields of exact agreement available for matching and their percentages of missing values are summarized in Multimedia Appendices 1-4 for the 4 use cases.

Within each run of the FS model, the estimate of block-specific prevalence under each missing treatment was used to classify record pairs as matches and nonmatches (see Classification of Record Pairs); the union of matches from all 5 blocks is the set of matches obtained.

### Evaluation of Record Linking Performance

To evaluate the accuracy of these matching models, we calculated the following metrics: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and $F_1$-score, as well as their respective 95% CIs based on 999 bootstrap samples. The abovementioned metrics were estimated from gold standard sets for which manual reviews established the true match status. Our primary evaluation metric was the $F_1$-score.

### Ethics Approval

This study was reviewed and approved by the Indiana University Institutional Review Board (IRB#: 1703755361).

## Results

The 4 use cases contained missing data to various extents. Notably, 45.8% of the 47,334,986 records in the INPC use case had no SSN, making it necessary to add other blocking schemes that do not rely on the SSN. For the NBS use case, the SSN is typically missing because infants do not receive an SSN for at least 2 to 6 weeks after birth and often later if parents do not initially request the identifier. When linking the other 2 use cases SSA and MCHD to INPC, due to the INPC data set missing SSN in 45.8% of its records, blocking on SSN alone yielded only 4547 out of 5950 (76%) and 1531 out of 5927 (26%) of true matches in SSA and MCHD, respectively, based on the manually reviewed subsets (Multimedia Appendices 5 and 6).

Additional blocking schemes are essential to increase match sensitivity. As the FS algorithm is performed using paired data per blocking scheme, its performance is directly affected by the extent of missing values in the agreement vectors obtained by comparing pairs of records within each block. We summarized the proportions of missing data in the 5 blocking schemes of each use case in Multimedia Appendices 1-4. The proportion of missing values per block ranges from 0% to 100%. Fields that were missing 100% within a blocking scheme contained no information and, therefore, were not considered further. The extent of missing values in a matching field does not necessarily negatively correlate with the discriminating power of the field.

Matching fields with even substantial missing values nonetheless proved to be useful in discriminating matches from nonmatches. For example, the agreement status of email address comparison is missing for 99% of record pairs in the DB-LN-MB-YB blocking scheme of the INPC use case; the m- and u-probabilities were estimated to be 0.01147 and 0.000204 under MAD and 0.3830 and 0.02553 under MAR, respectively. The large ratios of the m-probability over the u-probability in either case indicate the utility of email address in linkage. As another example, the agreement status of zip code comparison is also missing for 99% of record pairs in the FN-LN-MI-YB block of the SSA use case, and the m- and u-probabilities were estimated to be 0.02073 and $8.49\times10^{-7}$ under MAD and 0.7538 and 0.000137 under MAR, respectively. In both examples, the estimates of m- and u-probabilities are much larger under MAR than under MAD, suggesting that a downward bias might be incurred by artificially setting missing values to disagreement in the MAD approach.

The fields used by the final FS models, either expert-specified or data-driven per block per use case, are summarized in Table 3. Except for the SSA use case, where the number of fields that could be used for matching is limited, the number of data-driven fields is greater than the number of expert-specified fields for the remaining 3 use cases.

**Table 3.** Summary of modeling information by data use case and by blocking scheme.

| Data and block | Expert-specified fields[a] | Data-driven fields[a] |
|---|---|---|
| **INPC** | | |
| DB-LN-MB-YB[b] | MRN[c] FN[d] SEX[e] TEL[f] ADR[g] ZIP[h] SSN[i] | MRN FN SEX TEL ADR ZIP SSN *CITY EMAIL ETH MI NICK ST*[j] |
| DB-MB-YB-ZIP | MRN LN FN SEX TEL ADR SSN | MRN LN FN SEX TEL ADR SSN *CITY EMAIL ETH MI NICK ST* |
| FN-LN-YB | MRN SEX DB MB TEL ADR ZIP SSN | MRN SEX DB MB TEL ADR ZIP SSN *CITY EMAIL MI ST ETH NICK* |
| FN-TEL | MRN LN SEX DB MB YB ADR ZIP SSN | MRN LN SEX DB MB YB ADR ZIP SSN *CITY EMAIL ETH MI NICK ST* |
| SSN | MRN LN FN SEX DB MB YB TEL ADR ZIP | MRN LN FN SEX DB MB YB TEL ADR ZIP *CITY EMAIL ETH MI NICK ST* |
| **SSA[k]** | | |
| FN-LN-DB-MB-YB | SSN MI ZIP | SSN MI ZIP |
| FN-LN-MI-DB-MB | ZIP YB SSN | ZIP YB SSN |
| FN-LN-MI-YB | DB MB ZIP SSN | DB MB ZIP SSN |
| FN-LN-ZIP | MI DB MB YB SSN | MI DB MB YB SSN |
| SSN | LN FN MI DB MB YB ZIP | LN FN MI DB MB YB ZIP |
| **NBS[l]** | | |
| LN-FN | MRN SEX DB MB YB TEL ADR ZIP | MRN SEX[m] DB MB YB[m] TEL ADR ZIP *CITY DR_FN DR_LN MI NK_FN NK_LN* |
| MB-DB-ZIP | MRN LN FN SEX YB TEL ADR | MRN LN FN SEX YB TEL ADR *CITY DR_FN DR_LN ETH MI NK_FN NK_LN NICK* |
| MRN | LN FN SEX DB MB YB TEL ADR ZIP | LN FN SEX[m] DB MB YB TEL[m] ADR ZIP *CITY DR_FN DR_LN ETH MI NK_FN NK_LN ST* |
| NK_LN-NK_FN | MRN LN FN SEX DB MB YB TEL ADR ZIP | MRN LN[m] FN SEX DB MB YB TEL ADR ZIP *CITY DR_FN DR_LN ETH LN MI NICK ST* |
| TEL | MRN LN FN SEX DB MB YB ADR ZIP | MRN LN FN SEX DB MB YB[m] ADR ZIP *CITY DR_FN DR_LN ETH MI NK_FN NK_LN ST* |
| **MCHD[n]** | | |
| LN-FN | MRN SEX DB MB YB TEL ADR ZIP | MRN SEX[m] DB MB YB[m] TEL ADR ZIP *CITY DR_FN DR_LN MI NK_FN NK_LN* |
| MB-DB-ZIP | MRN LN FN SEX YB TEL ADR | MRN LN FN SEX YB TEL ADR *CITY DR_FN DR_LN ETH MI NK_FN NK_LN NICK* |
| MRN | LN FN SEX DB MB YB TEL ADR ZIP | LN FN SEX[m] DB MB YB TEL[m] ADR ZIP *CITY DR_FN DR_LN ETH MI NK_FN NK_LN ST* |
| NK_LN-NK_FN | MRN LN FN SEX DB MB YB TEL ADR ZIP | MRN LN[m] FN SEX DB MB YB TEL ADR ZIP *CITY DR_FN DR_LN ETH LN MI NICK ST* |
| TEL | MRN LN FN SEX DB MB YB ADR ZIP | MRN LN FN SEX DB MB YB[m] ADR ZIP *CITY DR_FN DR_LN ETH MI NK_FN NK_LN ST* |

[a] Columns “Expert-specified fields” and “Data-driven fields” display the fields used in the Fellegi-Sunter (FS) model.

[b] DB-LN-MB-YB: day, month, and year of birth and last name.

[c] MRN: medical record number.

[d] FN: first name.

[e] SEX: sex.

[f] TEL: telephone number.

[g] ADR: address.

[h] ZIP: zip code.

[i] SSN: Social Security number.

[j] Fields (italicized) selected only by data-driven methods.

[k] SSA: Social Security Administration.

[l] NBS: newborn screening.

[m] Fields not selected by the data-driven method but specified by experts.

[n] MCHD: Marion County Health Department.

The matching metrics of the 4 use cases evaluated on their respective ground truth sets of randomly selected and manually reviewed record pairs are displayed in Table 4. Multimedia Appendices 5-7. From Table 4, we observe the following:

1. MAR improves the *F1*-score in general, whether matching fields are expert-specified or data-driven; the improvement in the *F1*-score comes from improved sensitivity with comparable or better PPV. The largest improvement in the *F1*-score occurred in the NBS use case, 0.874 with MAR using data-driven fields compared with 0.837 with MAD using expert-specified fields.
2. MAD using expert-specified fields had higher *F1*-scores than *F1*-scores using data-driven fields (except for NBS). As the number of data-driven fields is usually greater than the number of expert-specified fields in a block, we hypothesized that the artificial correlations among the large number of data-driven fields induced by MAD adversely affect the match performance.
3. MAR coupled with data-driven fields yielded *F1*-scores comparable to or larger than those of MAR with expert-specified fields and larger than the F1-scores of MAD with both methods of field selection.

In the SSA use case, the $F_1$-scores of both methods were similar, 0.873 for MAD and 0.875 for MAR, with either expert-specified matching fields or data-driven matching fields, because both field-selection approaches selected the same set of matching fields. We examined the classification results within the classes of true matches and true nonmatches in the ground truth set, on whether the classified matches and nonmatches were similar or whether the 2 methods made different mistakes. From the diagonals in Table 5, we can see that the 2 methods produce roughly congruent classification results in the classes of true matches and true nonmatches. FS under MAR is slightly more sensitive than FS under MAD: 26 true matches that are misclassified as nonmatches by MAD are recovered as matches by MAR; only 3 nonmatches are misclassified as matches by MAR, but 1 nonmatch that is misclassified as a match by MAD is correctly classified as a nonmatch by MAR. In summary, the classification results are similar in the SSA use case; FS under MAR is slightly more sensitive than FS under MAD, while maintaining PPV, NPV, and specificity (Table 4).

The algorithms performed differently, partly because of the different data quality of the use cases. The $F_1$-score was 0.979 for INPC but only 0.874 for NBS (Table 4). First, INPC and NBS use cases have very different patterns of missing data across matching fields; for example, SSN is missing from 52.6% to 69.7% of record pairs (except for the SSN block, which by definition has no missing SSN) across the 5 blocks in INPC, whereas SSN is missing in more than 98% of record pairs in NBS (Multimedia Appendices 1 and 3). Second, the discriminating powers of the same fields were different in the 2 use cases. A field has high discriminating power if its agreement rate is high among matches and low among nonmatches; otherwise, the data quality is indicated as low. For example, the probabilities of agreement in the fields of the LN and FN in the same DB-MB-YB and ZIP blocking scheme for INPC and NBS show differential data quality: 94.45% of matches of INPC agree on the LN, while only 89% of matches of NBS do; on the other hand, only 0.45% of nonmatches of INPC agree on the LN but 2.70% of nonmatches of NBS do; 94.36% of matches of INPC agree on the FN, while only 66.99% of matches of NBS do; only 0.40% of nonmatches of INPC agree on the FN but 3.45% of nonmatches of NBS do (Multimedia Appendix 8). As the matching score is a summation of the ratios of the agreement probabilities of the matches versus the nonmatches on the log-scale across all matching fields, data quality directly affects matching performance.

**Table 4.** Matching results of the four use cases evaluated on their respective ground truth sets of random-selected and manually reviewed record pairs.

| Data | Value, N | Sensitivity (95% CI) | Specificity (95% CI) | Positive predictive value (95% CI) | Negative predictive value (95% CI) | $F_1$-score (95% CI) |
|---|---|---|---|---|---|---|
| **Expert-specified fields** | | | | | | |
| **INPC**[a] | | | | | | |
| MAD[b] | 15,000 | 0.962 (0.958-0.967) | 0.990 (0.987-0.992) | 0.990 (0.988-0.992) | 0.960 (0.955-0.964) | 0.976 (0.974-0.978) |
| MAR[c] | 15,000 | 0.970 (0.966-0.974) | 0.988 (0.986-0.991) | 0.989 (0.987-0.991) | 0.968 (0.964-0.972) | 0.980 (0.977-0.982) |
| **SSA**[d] | | | | | | |
| MAD | 16,500 | 0.781 (0.770-0.792) | 0.995 (0.994-0.996) | 0.989 (0.986-0.992) | 0.890 (0.884-0.895) | 0.873 (0.866-0.879) |
| MAR | 16,500 | 0.785 (0.775-0.796) | 0.995 (0.993-0.996) | 0.989 (0.985-0.991) | 0.892 (0.886-0.897) | 0.875 (0.869-0.882) |
| **NBS**[e] | | | | | | |
| MAD | 15,000 | 0.795 (0.786-0.804) | 0.881 (0.874-0.889) | 0.883 (0.876-0.891) | 0.791 (0.782-0.801) | 0.837 (0.830-0.843) |
| MAR | 15,000 | 0.860 (0.852-0.868) | 0.873 (0.865-0.881) | 0.885 (0.877-0.892) | 0.846 (0.838-0.855) | 0.872 (0.866-0.878) |
| **MCHD**[f] | | | | | | |
| MAD | 15,500 | 0.944 (0.937-0.949) | 0.989 (0.987-0.991) | 0.982 (0.979-0.986) | 0.966 (0.962-0.969) | 0.963 (0.959-0.966) |
| MAR | 15,500 | 0.946 (0.940-0.952) | 0.988 (0.986-0.990) | 0.980 (0.976-0.983) | 0.967 (0.964-0.971) | 0.963 (0.959-0.966) |
| **Data-driven fields** | | | | | | |
| **INPC** | | | | | | |
| MAD | 15,000 | 0.579 (0.568-0.590) | 0.988 (0.986-0.991) | 0.982 (0.978-0.985) | 0.682 (0.672-0.690) | 0.729 (0.719-0.737) |
| MAR | 15,000 | 0.970 (0.966-0.974) | 0.987 (0.984-0.989) | 0.988 (0.985-0.990) | 0.968 (0.964-0.972) | 0.979 (0.976-0.981) |
| **SSA** | | | | | | |
| MAD | 16,500 | 0.781 (0.770-0.792) | 0.995 (0.994-0.996) | 0.989 (0.986-0.992) | 0.890 (0.884-0.895) | 0.873 (0.866-0.879) |
| MAR | 16,500 | 0.785 (0.775-0.796) | 0.995 (0.993-0.996) | 0.989 (0.985-0.991) | 0.892 (0.886-0.897) | 0.875 (0.869-0.882) |
| **NBS** | | | | | | |
| MAD | 15,000 | 0.813 (0.805-0.822) | 0.875 (0.867-0.883) | 0.880 (0.873-0.888) | 0.805 (0.796-0.814) | 0.845 (0.839-0.852) |
| MAR | 15,000 | 0.865 (0.858-0.873) | 0.870 (0.863-0.878) | 0.883 (0.876-0.890) | 0.851 (0.842-0.859) | 0.874 (0.868-0.880) |
| **MCHD** | | | | | | |
| MAD | 15,500 | 0.635 (0.622-0.648) | 0.970 (0.967-0.974) | 0.929 (0.921-0.937) | 0.811 (0.804-0.818) | 0.754 (0.745-0.764) |
| MAR | 15,500 | 0.954 (0.948-0.959) | 0.988 (0.985-0.990) | 0.979 (0.976-0.983) | 0.972 (0.968-0.975) | 0.967 (0.963-0.970) |

[a] INPC: Indiana Network for Patient Care.

[b] MAD: missing as disagreement.

[c] MAR: missing at random.

[d] SSA: Social Security Administration.

[e] NBS: newborn screening.

[f] MCHD: Marion County Health Department.

**Table 5.** Cross-tabulation of ground truth and classification results by the Fellegi-Sunter model under missing as disagreement (MAD) and missing at random (MAR) for the Social Security Administration use case.

| MAD | MAR | | Values, N |
|---|---|---|---|
| | Nonmatch | Match | |
| **Man_Rev_Status: matches** | | | |
| Nonmatch | 1277 | 26 | 1303 |
| Match | 0 | 4647 | 4647 |
| Value, N | 1277 | 4673 | 5950 |
| **Man_Rev_Status: nonmatches** | | | |
| Nonmatch | 10,495 | 3 | 10,498 |
| Match | 1 | 51 | 52 |
| Value, N | 10,496 | 54 | 10,550 |

## Discussion

The US health care system will likely not have a unique and universal patient ID in the near future, so innovations such as incorporating missing data under MAR and data-driven field selection in the linkage algorithms are necessary to optimize existing methods to ensure accurate patient identity and support patient safety. Our findings are important because they demonstrate improvements in linkage performance among 4 different but representative use cases. Our HIE-based patient-matching laboratory has experience matching clinical data from heterogeneous sources, including hospitals (inpatient and emergency departments) [21,22], ambulatory care settings [23], public health syndrome as captured by surveillance systems [24], electronic laboratory reporting in case detection systems [25], and NBS data [26]. Thus, we can specifically measure variations in patient-matching performance across different use cases for the same patient-matching approaches. Through this work, we have shown that the performance of the 2 enhancements in patient-matching algorithms may have additive values across different clinical use cases.

Although the assumption of missing at random is not verifiable, the success of the FS algorithm coupled with MAR in our four different use cases indicates that missing at random is a reasonable assumption. As MCAR is a special case of MAR, our algorithm works when data are MCAR. These results will inform future research and development in patient-matching spaces.

Furthermore, the superior performance observed with MAR using data-driven fields over other combinations in the 2×2 design and four use cases suggests its potential value for incorporation into privacy-preserving record linkage (PPRL) methods. In PPRL, to preserve privacy, fields can be tokenized (eg, using bigrams) into smaller parts and compared [27]. As the tokens do not reveal the actual nature and content of the field, an expert cannot specify matching fields as they can with fields such as names, DB, SSN, etc. PPRL is a scenario in which data-driven field selection coupled with MAR in the FS model appears useful.

Finally, many data-driven fields may lead to model overfitting, which is a prominent cause of the poor performance of machine learning algorithms. In many applications in medical research using latent class models, many covariates are available, and the number of covariates overwhelms the number of observations. This is the main motivation for most of the variable selection literature to identify a subset of variables to (1) estimate the association between the covariates and the response variable and (2) obtain a parsimonious model that describes the covariates and the response variable [28]. However, in record linkage, the primary concern is not estimation but rather the prediction of the *unknown* response variable—the class label of match or nonmatch, model parsimony is irrelevant. Sample sizes in record linkage are generally large: in our four use cases, even the sample size of the smaller class of our smallest use case, MCHD, is approximately 180,035 (Tables 1 and 2). Hence, the number of matching fields used relative to the overall samples and the within-class sample sizes were small, without concern for overfitting. The FS model for record linkage is an unsupervised algorithm in that the response variable indicates whether a pair of records belonging to the same entity is unknown and is to be inferred. As such, the veracity of any unsupervised classification algorithm applied to a set of record pairs is tested on a representative ground truth set in which the class labels are obtained through manual reviews. As described in the Methods section, we created a gold standard set of randomly selected record pairs for each of the 4 use cases. Most importantly, the match status of those record pairs in the gold standard sets was not used in the FS model fitting process (including field selection). As the data-driven fields used in the FS model under MAR uniformly perform the best for all four use cases, overfitting is not a concern because overfitting tends to hurt performance.

While we strive to generate results that are applicable to the broadest possible audience using a health informatics research laboratory that captures a diverse set of data elements with varying data characteristics, we cannot assure generalizability with complete certainty. If our data are not representative of other health systems, then our linkage results may not be applicable. If the missing data mechanism is not MAR or MCAR (eg, if the missingness of a data element is related to its value), our algorithm will likely not work. Before applying our methods to a data environment with missing data, we recommend creating a ground truth set of randomly selected record pairs whose match status is manually reviewed to determine whether our methods are applicable to a specific data environment.

Finally, our results suggest that accommodating missingness in patient-matching algorithms can improve accuracy. While the FS model is widely used, different FS implementations and completely different models (eg, decision trees or boosting algorithms) may exhibit a greater or lesser effect. We will explore the potential of these machine learning tools in our future work.

In summary, the combination of data-driven matching field selection and MAR methods produced the best overall performance for four real-world matching use cases. The MAR method maintained or improved $F_1$-scores regardless of whether matching fields were expert-specified or data-driven, suggesting that MAR is a reasonable assumption for patient-record linkage in real-world settings. As the implementation of MAR requires minimal effort and improves or maintains linkage accuracy, we advocate using this approach over MAD in record linkage, provided that adequate evaluation using manually reviewed data is performed to ensure method generalizability to a specific data environment. These methods can be useful for PPRL, where expert field selection may not be possible.

## Acknowledgments

This research was supported by grants (Agency for Healthcare Research and Quality) and (The Patient-Centered Outcomes Research Institute).

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Table S1 Proportion of missing values by matching field in the Indiana Network for Patient Care (INPC) use case. For each blocking scheme (column) the unshaded fields are used for matching in the final Fellegi-Sunter (FS) model for that block in the data-driven approach.

[DOCX File , 38 KB - jmir_v24i9e33775_app1.docx ]

Multimedia Appendix 2

Table S2 Proportion of missing values by field in the Social Security Administration (SSA) use case. For each blocking scheme (column) the unshaded fields are used for matching in the final Fellegi-Sunter (FS) model for that block in the data-driven approach.

[DOCX File , 39 KB - jmir_v24i9e33775_app2.docx ]

Multimedia Appendix 3

Table S3 Proportion of missing values by field in the newborn screening (NBS) use case. For each blocking scheme (column) the unshaded fields are used for matching in the final Fellegi-Sunter (FS) model for that block in the data-driven approach.

[DOCX File , 39 KB - jmir_v24i9e33775_app3.docx ]

Multimedia Appendix 4

Table S4 Proportion of missing values by field in the Marion County Health Department (MCHD) use case. For each blocking scheme (column) the unshaded fields are used for matching in the final Fellegi-Sunter (FS) model for that block in the data-driven approach.

[DOCX File , 39 KB - jmir_v24i9e33775_app4.docx ]

Multimedia Appendix 5

Table S5 Matching results of the SSA use case evaluated on a set of 16,500 randomly selected and manually reviewed record pairs. The first two rows are the overall results combined from all blocks on the manually reviewed sample, with the first row for MAD (missing as disagreement) and the second row for MAR (missing at random). Every subsequent two rows pertain to a specific block, with the first containing the results of MAD and the 2nd row the results of MAR. Columns N, SEN, SPE, PPV, NPV and F1 are the total number of manually reviewed record pairs, sensitivity, specificity, positive predictive value, negative predictive value and F-score.

[DOCX File , 42 KB - jmir_v24i9e33775_app5.docx ]

Multimedia Appendix 6

Table S6 Matching results of the newborn screening (NBS) use case evaluated on a set of 15,000 randomly selected and manually reviewed record pairs. The first two rows are the overall results combined from all blocks on the manually reviewed sample, with the first row for MAD (missing as disagreement) and the second row for MAR (missing at random). Every subsequent two rows pertain to a specific block, with the first containing the results of MAD and the 2nd row the results of MAR. Columns N, SEN, SPE, PPV, NPV, and F1 are the total number of manually reviewed record pairs, sensitivity, specificity, positive predictive value, negative predictive value and F-score.
[DOCX File , 42 KB - jmir_v24i9e33775_app6.docx ]

---

Multimedia Appendix 7
Table S7 Matching results of the Marion County Health Department (MCHD) use case evaluated on a set of 15,500 randomly selected and manually reviewed record pairs. The first two rows are the overall results combined from all blocks on the manually reviewed sample, with the first row for MAD (missing as disagreement) and the second row for MAR (missing at random). Every subsequent two rows pertain to a specific block, with the first containing the results of MAD and the 2nd row the results of MAR. Columns N, SEN, SPE, PPV, NPV, and F1 are the total number of manually reviewed record pairs, sensitivity, specificity, positive predictive value, negative predictive value and F-score.
[DOCX File , 42 KB - jmir_v24i9e33775_app7.docx ]

---

Multimedia Appendix 8
Table S8 Data quality of fields of last name and first name in the DOB-ZIP block of the Indiana Network for Patient Care (INPC) and newborn screening (NBS) use cases.
[DOCX File , 36 KB - jmir_v24i9e33775_app8.docx ]

---

## References

1. Fellegi IP, Sunter AB. A theory for record linkage. J Am Stat Assoc 1969 Dec;64(328):1183-1210. [doi: 10.1080/01621459.1969.10501049]
2. Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a probabilistic record linkage technique without human review. AMIA Annu Symp Proc 2003:259-263 [FREE Full text] [Medline: 14728174]
3. Culbertson A, Goel S, Madden M, Safaeinili N, Jackson K, Carton T, et al. The building blocks of inter-operability. Appl Clin Inform 2017 Dec 21;08(02):322-336. [doi: 10.4338/aci-2016-11-ra-0196]
4. Tromp M, Ravelli AC, Méray N, Reitsma JB, Bonsel GJ. An efficient validation method of probabilistic record linkage including readmissions and twins. Methods Inf Med 2008 Jan 18;47(4):356-363. [doi: 10.3414/me0489] [Medline: 18690369]
5. Tromp M, Reitsma JB, Ravelli AC, Méray N, Bonsel GJ. Record linkage: making the most out of errors in linking variables. AMIA Annu Symp Proc 2006:779-783 [FREE Full text] [Medline: 17238447]
6. Daggy JK, Xu H, Hui SL, Gamache RE, Grannis SJ. A practical approach for incorporating dependence among fields in probabilistic record linkage. BMC Med Inform Decis Mak 2013 Aug 30;13(1):97 [FREE Full text] [doi: 10.1186/1472-6947-13-97] [Medline: 24001000]
7. Brown AP, Randall SM, Ferrante AM, Semmens JB, Boyd JH. Estimating parameters for probabilistic linkage of privacy-preserved datasets. BMC Med Res Methodol 2017 Jul 10;17(1):95 [FREE Full text] [doi: 10.1186/s12874-017-0370-0] [Medline: 28693507]
8. Allison P. Missing Data. Thousand Oaks CA: Sage; 2001.
9. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006 Oct;59(10):1087-1091. [doi: 10.1016/j.jclinepi.2006.01.014] [Medline: 16980149]
10. Jones MP. Indicator and stratification methods for missing explanatory variables in multiple linear regression. J Am Stat Assoc 1996 Mar;91(433):222-230. [doi: 10.1080/01621459.1996.10476680]
11. Knol MJ, Janssen KJ, Donders AR, Egberts AC, Heerdink ER, Grobbee DE, et al. Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. J Clin Epidemiol 2010 Jul;63(7):728-736. [doi: 10.1016/j.jclinepi.2009.08.028] [Medline: 20346625]
12. Vach W, Blettner M. Missing data in epidemiologic studies. In: Encyclopedia of Biostatistics. Hoboken, New Jersey, United States: Wiley; 2005.
13. Vink G, Frank LE, Pannekoek J, van Buuren S. Predictive mean matching imputation of semicontinuous variables. Statistica Neerlandica 2014 Jan 28;68(1):61-90. [doi: 10.1111/stan.12023]
14. ENAMORADO T, FIFIELD B, IMAI K. Using a probabilistic model to assist merging of large-scale administrative records. Am Polit Sci Rev 2019 Jan 02;113(2):353-371. [doi: 10.1017/s0003055418000783]
15. Xu H, Li X, Shen C, Hui SL, Grannis S. Incorporating conditional dependence in latent class models for probabilistic record linkage: does it matter? Ann Appl Stat 2019 Sep 1;13(3):1753-1790. [doi: 10.1214/19-aoas1256]
16. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the algorithm. J Royal Statis Soc Series B (Methodological) 2018 Dec 05;39(1):1-22. [doi: 10.1111/j.2517-6161.1977.tb01600.x]
17. Little R, Rubin D. Statistical Analysis with Missing Data. Hoboken, New Jersey, United States: Wiley; 2014.

18. Steorts R, Ventura S, Sadinle M, Fienberg S. A comparison of blocking methods for record linkage. In: Privacy in Statistical Databases. Cham: Springer International Publishing; 2014.
19. Jain A, Flynn P, Ross A. Handbook of Biometrics. Cham: Springer; 2008.
20. Prabhakar S, Pankanti S, Jain A. Biometric recognition: security and privacy concerns. IEEE Secur Privacy 2003 Mar;1(2):33-42. [doi: 10.1109/msecp.2003.1193209]
21. Finnell J, Overhage J, Grannis S. All health care is not local: an evaluation of the distribution of Emergency Department care delivered in Indiana. AMIA Annu Symp Proc 2011;2011:409-416 [FREE Full text] [Medline: 22195094]
22. Wu J, Grannis SJ, Xu H, Finnell JT. A practical method for predicting frequent use of emergency department care using routinely available electronic registration data. BMC Emerg Med 2016 Feb 09;16(1):12 [FREE Full text] [doi: 10.1186/s12873-016-0076-3] [Medline: 26860825]
23. Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. Am J Public Health 2008 Feb;98(2):344-350. [doi: 10.2105/ajph.2006.092700]
24. Grannis S, Wade M, Gibson J, Overhage JM. The Indiana public health emergency surveillance system: ongoing progress, early findings, and future directions. AMIA Annu Symp Proc 2006:304-308 [FREE Full text] [Medline: 17238352]
25. Gichoya J, Gamache RE, Vreeman DJ, Dixon BE, Finnell JT, Grannis S. An evaluation of the rates of repeat notifiable disease reporting and patient crossover using a health information exchange-based automated electronic laboratory reporting system. AMIA Annu Symp Proc 2012;2012:1229-1236 [FREE Full text] [Medline: 23304400]
26. Zhu VJ, Overhage MJ, Egg J, Downs SM, Grannis SJ. An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. J Am Medical Inform Assoc 2009 Sep 01;16(5):738-745. [doi: 10.1197/jamia.m3186]
27. Schnell R, Bachteler T, Reiher J. Privacy-preserving record linkage using Bloom filters. BMC Med Inform Decis Mak 2009 Aug 25;9(1):41 [FREE Full text] [doi: 10.1186/1472-6947-9-41] [Medline: 19706187]
28. Khalili A, Chen J. Variable selection in finite mixture of regression models. J Am Statistical Assoc 2012 Jan 01;102(479):1025-1038. [doi: 10.1198/016214507000000590]

## Abbreviations

**CITY:** city
**DB:** day of birth
**EM:** Expectation-Maximization
**EMR:** electronic medical record
**ETH:** ethnicity
**FN:** first name
**FS:** Fellegi-Sunter
**HIE:** health information exchange
**HL7:** Health Level Seven International
**INPC:** Indiana Network for Patient Care
**LN:** last name
**LN-FN:** last name and first name
**MAR:** missing at random
**MB:** month of birth
**MCAR:** missing completely at random
**MCHD:** Marion County Health Department
**MI:** middle initial
**MNAR:** missing not at random
**MRN:** medical record number
**NBS:** newborn screening
**NPV:** negative predictive value
**PPRL:** privacy-preserving record linkage
**PPV:** positive predictive value
**SSA:** Social Security Administration
**SSN:** Social Security number
**YB:** year of birth

*Edited by G Eysenbach; submitted 22.09.21; peer-reviewed by V Rajan, Y Chu, T Lefèvre; comments to author 29.03.22; revised version received 16.05.22; accepted 28.07.22; published 29.09.22.*

*Please cite as:*
*Li X, Xu H, Grannis S*
*The Data-Adaptive Fellegi-Sunter Model for Probabilistic Record Linkage: Algorithm Development and Validation for Incorporating Missing Data and Field Selection*
*J Med Internet Res 2022;24(9):e33775*
*URL: https://www.jmir.org/2022/9/e33775*
*doi:10.2196/33775*
*PMID:36173664*

Original Paper

# Laypeople’s Online Health Information Search Strategies and Use for Health-Related Problems: Cross-sectional Survey

Yen-Lin Chiu[1], PhD; Chin-Chung Tsai[2,3], PhD; Jyh-Chong Liang[2,3], PhD

[1]Graduate Institute of Medical Education and Bioethics, National Taiwan University College of Medicine, Taipei, Taiwan
[2]Program of Learning Sciences, National Taiwan Normal University, Taipei, Taiwan
[3]Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei, Taiwan

**Corresponding Author:**
Jyh-Chong Liang, PhD
Program of Learning Sciences
National Taiwan Normal University
No. 162, Section 1
Heping East Road
Taipei, 106
Taiwan
Phone: 886 2 7749 5999
Email: aljc@ntnu.edu.tw

## Abstract

**Background:** With the increase in the use of the internet to search for health information about health-related problems, there is a need for health care professionals to better understand how their patients search for and use the online health information that may influence their medical decision making.

**Objective:** The aims of this study are to explore laypeople’s online health information search strategies and examine the relationships between their search strategies and utilization behavior of online health information.

**Methods:** Two scales, namely match and elaboration, were used to measure patients’ basic search strategies (ie, simple approach) and advanced search strategies (ie, integrative approach), respectively. In addition, the consultation scale was used to evaluate the participants’ use of online health information to consult doctors and others. A total of 253 outpatients without university education were purposely selected and surveyed. The participants were outpatients at a university-affiliated teaching hospital. Partial least squares-structural equation modeling (PLS-SEM) was performed to analyze the measurement model to specify the measurement validation. In addition, the structure model of PLS-SEM was evaluated to examine the path correlations between variables and to execute interaction effect and curvilinear relationship analyses.

**Results:** The results of the path correlation analysis by PLS-SEM showed that both elaboration strategy (path coefficient=0.55, $P<.001$) and match strategy (path coefficient=0.36, $P<.001$) were positively correlated with consultation on online health information with doctors and others. In addition, interaction effect and curvilinear relationship analyses indicated that there was a significant interaction effect between elaboration and match on consultation (path coefficient=–0.34, $P<.001$) and a significant curvilinear relationship between match and consultation (path coefficient=–0.09, $P=.046$).

**Conclusions:** Increasing patients’ exposure to online health information through both a simple search approach (ie, match strategy) and a complex search approach (ie, elaboration strategy) may lead them to appropriately use the information to consult doctors and others. However, the results of interaction effect and curvilinear relationship analyses highlighted the essential role of the elaboration strategy to properly locate, evaluate, and apply online health information. The findings of this study may help health care professionals better understand how to communicate with their patients through the health information on the internet.

***(J Med Internet Res 2022;24(9):e29609)*** doi:10.2196/29609

**KEYWORDS**

decision making; eHealth literacy; information search strategy; internet; patient; information-seeking behavior; laypeople; online health information; patient communication

## Introduction

### Background

With its convenient and widespread access to abundant information, the internet has become the major source for patients and the general population to retrieve health information [1]. As reported by the Pew Research Center, approximately 80% of American internet users search the internet for online health information [2]. In Taiwan, it is estimated that 83.4% of residents aged 12 years and above have internet experience [3]. As reported by the Taiwan National Health Interview Survey, 1766 (64.4%) of the 2741 surveyed individuals used the internet to search for online health information or services [4]. The issues regarding online health information–seeking behaviors of patients have attracted a great deal of attention, since the health information located on the internet obviously influences patients' medical decision making [5-7].

Having better access to health information on the internet provides internet users with more possibilities to actively manage their own health and medical utilization behaviors [8]. The internet is regarded as a powerful and influential tool through which retrieving online health information may benefit patients' empowerment, well-being health change, and healthier behaviors [9,10]. Compared to infrequent users, frequent internet users prefer more health-related information and decision making and the internet enables them to make more informed medical decisions [11]. In addition to medical decision making, online health information influences patients' communication with physicians [12]. The result of a systematic review study showed that online health information improves patient-physician relationships as patients gain better access to online health information and discuss it with their physicians [13].

Despite the use of the internet to search for health information making internet users more knowledgeable, patients seldom discuss the information they find on the web with their doctors [14]. The credibility of the diverse range of health information on the internet has been of great concern, as its inappropriate use may be potentially harmful to patients' health and waste medical resources [15,16]. In sum, online health information without verification by experts could generate misinformation and inappropriate health behaviors and hinder the physician-patient relationship [16,17].

According to systematic review studies, it has been concluded that the overall quality of online health information remains problematic and should be considered [18,19]. However, a high percentage (77%) of internet users tend to search for health information through search engines due to the decentralized nature of the internet [2]. An observational study on health information–seeking behaviors showed a high tendency of using search engines to look for health information [20]. As indicated by an experiment, the most popular method for seeking health information was to rely on the results of only 1 search engine page and to use unaccredited information to answer health questions without comparing and justifying them with other sources [5]. Because of the high heterogeneity of online health information sources, rather than merely relying on the first few results provided by the search engine, the need to check certain information against other information sources while searching for online information about medical problems has been strongly recommended [21,22]. The impacts of internet search strategies on information retrieval and how patients use online health information have been of great concern; however, the information search strategies for health-related problems have seldom been studied [7,23]. As suggested, there is a need to conduct more in-depth surveys to better understand how online health information–seeking behaviors influence the use of information in health-related decision making [24].

With respect to the role of education in the use of the internet for health information searching, it was reported that higher education is significantly connected to a higher probability of using the internet as the first source of health information [1]. The results of a population-based survey showed that respondents with lower education levels less frequently access health information from internet websites, while individuals with university or higher education more frequently search the internet for health information [10,25]. Regarding the use of online health information, the role of education has been linked to the use of credible information in health-related decision making [26]. While looking for health care providers to solve their medical problems, adult individuals with less formal education are less likely to use online resources to consult online rankings and reviews of doctors, hospitals, drugs, and medical treatments [27]. In a study on health information–related seeking and sharing behaviors among baby boomers and older adults, the results showed that college graduates and postgraduates are more likely than non-high-school graduates to seek and share health information over the internet [24].

According to a systematic review on studies that measured online health information usage, it was found that online health information can support desired health decisions, including increasing professional visits, asking questions during medical consultations, and adhering to physicians' advice [28]. This review paper suggested that future studies strictly validate instruments for investigating online health information–seeking behaviors and carefully examine their impacts on health decisions. Using the concept analysis methodology, 1 study conducted a systematic review on the past 10 years of research to analyze the concept of health information–seeking behavior. The concept analysis results pointed out that the internet has become a common and preferred channel for retrieving health information. In addition to the importance of investigating how individuals from different communities seek information on the internet, the results of this study highlighted the lack of scales that can further measure and understand health information–seeking behaviors. They also concluded that there is a need to advance individuals' ability to adequately acquire online health information and properly act on the acquired information to make health decisions [29].

### Research Purposes

Low levels of education have been correlated with undesirable online health information–seeking behaviors [24,27]. Research on the health information–seeking behaviors of the general population without a university education has been an issue of concern [22]. In addition, it was indicated that laypeople without a university background may receive less training in information search strategies and have difficulties searching for health information on the internet [5]. However, it was argued that either simple or complex search strategies would benefit general health information seekers to gather useful health information [5,20]. However, it was reported that those with low educational levels may not benefit from online health information, since they do not access alternative health information from health care providers [30]. As suggested, it is a major topic to explore how laypeople conduct either basic or advanced search strategies to obtain online health information to investigate and solve their health problems [20,31].

Therefore, this study examined in which manners patients without a university degree search for health information over the internet and how they use that information to answer their health-related questions. Thus, the correlations between patients' health information search strategies and utilization behaviors were explored. Since the variables of search frequency, age, and sex were regarded as influential demographics in patients' health information–seeking behaviors as well as doctor-patient consultations [32,33], these variables were also measured and recruited in the analyses and treated as control variables. Based on the aforementioned objectives, this study aimed to examine the following research questions:

- Question 1: Are there correlations between laypeople's health information search strategies and their health information utilization?
- Question 2: Are there interaction relationships between health information search strategies and health information utilization?
- Question 3: Are there curvilinear relationships between health information search strategies and health information utilization?

# Methods

## Recruitment

To examine laypeople with a low-level education background, a probable sample of outpatients without university education was purposefully selected and surveyed in a large-scale, university-affiliated teaching hospital. The criterion for recruiting participants was having experience of searching for online health information. All the participants surveyed were patients who visited an outpatient clinic for health-related problems and consulted a doctor about their problems. All the participants voluntarily participated in this study by responding to the survey. Informed consent for the survey was obtained from individual participants. In addition, their privacy has been strictly protected.

## Instruments

### *Procedure for Developing and Validating the Measurements*

According to the process suggested, the measurement development of this study was conducted in several steps involving theoretical and practical considerations [34]. With respect to the theoretical aspect, 2 measurements, the Information Commitment Survey (ICS) and Online Health Information Utilization (OHIU), were adopted from previous works that have involved clearly conceptual definitions and a theoretical basis for these measurements [7,35]. After receiving permission from the corresponding authors of these studies, the Chinese versions of the ICS and the OHIU were obtained and used in this study. Next, the wording of the items relating to the elaboration strategy, match strategy, and consultation were carefully modified to assess individuals' opinions on searching for and using online health information. To ensure content validity, we requested 2 medical experts and 1 expert in information science to evaluate the correspondence between the individual item and its theoretical construct. In addition, we purposively recruited 10 representative participants in a pilot test to subjectively check whether the wording and readability of the ICS and OHIU were appropriate. Finally, we conducted partial least squares-structural equation modeling (PLS-SEM) to analyze the measurement model and examine the reliability, discriminant validity, and convergent validity of the measurements.

### *Demographic Variables*

Demographic variables, including age, sex, and search frequency, were measured and recruited in the statistical analyses. Age was the participants' actual age. For sex, males were coded as 1, while females were coded as 2. The search frequency, that is, the patients' frequency of using the internet to search for health information for health-related problems, was measured with a 6-point scale ranging from 1 (rarely) to 6 (always).

### *Information Commitment Survey*

Two constructs retrieved from the ICS signified web users' information search strategies, namely the elaboration strategy and the match strategy [34,35]. These 2 constructs were modified and used to assess patients' online information search strategies for answering their health-related questions. These measurements were evaluated with a 6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree), indicating participants' opinions on each item of the search strategy. The details of the elaboration and match strategies are as follows.

- Elaboration as a search strategy (elaboration): evaluating the extent to which web users have metacognitive thinking and integrate information from diverse websites to find the best solution to fulfill their purposes. Example item: I can integrate the information retrieved from various websites.
- Match as a search strategy (match): assessing the extent to which web users wish to find a few websites containing fruitful and relevant information to match their searching purposes. Example item: I wish to find a single website containing the most fruitful information.

### *Online Health Information Utilization*

The online health information consulting scale, named consultation, which is a subscale of the OHIU questionnaire, presented patients' behaviors of using the health-related information retrieved from the web to consult doctors, experts, and relatives [7]. The items of consultation were measured with a 6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree), indicating participants' opinions on their consulting behaviors. The definition of consultation is as follows:

- Consulting scale (consultation): measuring the extent to which patients consult others about the online health information they retrieve. Example item: I will discuss with a physician the issues regarding the medical information retrieved on the internet.

### Data Analysis

Statistical software packages for social science SPSS Statistics version 22 (IBM Corp) and SmartPLS3 (SmartPLS GmbH, Germany) were used to conduct statistical analyses. Using partial least squares-structural equation modeling (PLS-SEM) analysis, the measurement model of 2 instruments and the structural model of the research hypotheses were examined based on the 2-stage procedure recommended by Hair et al [36]. The statistical software SmartPLS3 was used to execute the PLS-SEM procedure. First, this study evaluated the reliability and validity of the ICS and OHIU instruments, including factor loadings, composite reliability (CR), average variance explained (AVE), and the Fornell-Lacker criterion [37]. Next, we executed path correlation analysis to examine the relationships among the participants' age, sex, search frequency, elaboration, match, consultation, moderating term of elaboration and match, quadratic term of elaboration, and quadratic term of match. *P* values less than .05 indicated significant loadings and significant correlations between variables. Moreover, CR values greater than 0.7 and AVE values greater than the threshold value of 0.5 were considered as having adequate construct reliability and acceptable convergent validity, respectively [38].

### Ethical Considerations

This study was exempt from Institutional Review Board oversight in accordance with Article 5 of the Human Subjects Research Act of the Ministry of Health and Welfare, Republic of China (Taiwan) [39], and the "Scope of Human Research Cases Exempt from Ethical Review Board Review" announced by Ministry of Health and Welfare, Taiwan on 5 July 2012, pursuant to Wei-Shu-Yi-Zi (#1010265075) [40].

The research involved the use of questionnaires and survey procedures and was conducted in a public setting. The information obtained was recorded in such a manner that human subjects cannot be identified, directly or through identifiers linked to the subjects. Informed consent was obtained from all participants involved in the study, and the participants were subjected to no medical interactions or interventions other than ongoing usual care. The study was also conducted in accordance with the ethics standards required by the Declaration of Helsinki issued in 2013.

## Results

### Participants

A sample of 253 outpatients without a university academic degree was recruited for this study. The participants included 134 (53%) males and 119 (47%) females, who were outpatients at a university-affiliated teaching hospital in the northern area of Taiwan. Their average age was 45.73 (range 30-69) years.

### Results of Correlation Analysis

Table 1 provides the means and SDs of the variables and the Pearson correlation coefficients between them. As shown in Table 1, elaboration was linked to age (*r*=0.17, *P*<.01) and search frequency (*r*=0.24, *P*<.001) with positive correlation coefficients. In addition, both elaboration strategy (*r*=0.55, *P*<.001) and match strategy (*r*=0.31, *P*<.001) were positively correlated with consultation. That is, patients with high intent to conduct elaboration and match searches were more likely to consult others about the online health information they retrieved.

**Table 1.** Means (SDs) and correlations of variables.

| Variables | Mean (SD) | Correlation | | | |
|---|---|---|---|---|---|
| | | Age | Search frequency | Elaboration | Match |
| Age | 45.73 (7.70) | N/A[a] | N/A | N/A | N/A |
| Search frequency | 3.41 (1.07) | –0.07 | N/A | N/A | N/A |
| Elaboration | 4.68 (0.74) | 0.17[b] | 0.24[c] | N/A | N/A |
| Match | 4.36 (0.84) | 0.11 | –0.06 | 0.09 | N/A |
| Consultation | 4.38 (1.07) | 0.18[b] | 0.09 | 0.55[c] | 0.31[c] |

[a] N/A: not applicable.

[b] *P*<.01.

[c] *P*<.001.

### PLS-SEM Analysis of the Measurement Model

PLS-SEM analysis of measurement model showed that the 9 items of 3 factors (elaboration, match and, consultation) had significant and satisfactory factor loadings ranging from 0.60 to 0.92. The CR value of each construct was fairly good, ranging from 0.81 to 0.89. Moreover, the AVE values were larger than the threshold value of 0.5, ranging from 0.59 to 0.74, showing acceptable convergent validity [41]. Based on the Fornell-Lacker criterion, the square root of the AVE for each factor (ranging from 0.77 to 0.86) was higher than the corresponding interfactor correlations (ranging from 0.09 to 0.55), suggesting reasonable discriminant validity [37]. For details of the measurement model analysis, please refer to Multimedia Appendix 1.

### Path Correlation Analysis of the Structural Model

Combined with main variables, demographic variables, moderating term, and 2 quadradic terms, path correlation analysis was performed using SmartPLS3. The main variables containing elaboration, match, consultation, and demographic variables, including sex, age, and search frequency, were involved in the structural model to evaluate the path coefficients between the variables. To further examine the nonlinear effects of elaboration and match on consultation, following the procedure suggested, we used the 2-stage approach to create a moderating term (interaction effect between elaboration and match) and 2 quadratic terms (quadratic effects of elaboration and match) on the basis of standardized data [42].

Figure 1 presents the path coefficients of the structural model. The elaboration (path coefficient=0.55, *P*<.001) and match (path coefficient=0.36, *P*<.001) showed positive correlations between consultation, while the moderating term (path coefficient=–0.34, *P*<.001) and quadratic terms of match (path coefficient=–0.09, *P*=.05) showed negative correlations with consultation. Regarding demographics, sex, age, and search frequency did not have significant correlation with consultation. Overall, the $R^2$ value for consultation was 0.49, while the adjusted $R^2$ value was 0.47. In addition, the $f^2$ values of elaboration, match, moderating term, and quadratic terms ranging from 0.18 to 0.43 were higher than 0.025, showing large effects of independent variables [42]. Moreover, the values of the variance inflation factor (VIF) for independent variables ranged from 1 to 2.84, indicating that there was no problem of collinearity [41].

To further illustrate the curvilinear relationship of match with consultation, we used the means of latent variables calculated by PLS to estimate the quadratic equation of consultation on match. The scatter plot with its trend curve is plotted in Figure 2. As presented, the coefficient of x was positive, while the coefficient of $X^2$ was negative, indicating a concave downward relationship between match and consultation. That is, an increase in match had an initial positive effect on consultation, but the effect became weaker and even changed direction when match reached a high level, suggesting that match has a decrement of positive effect on consultation.

**Figure 1.** Path correlations of the structure model. *P<.05, ***P<.001

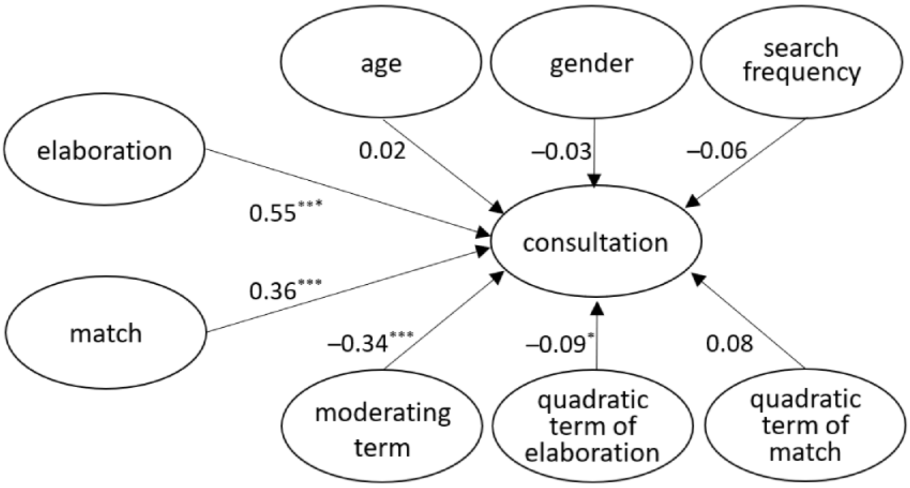
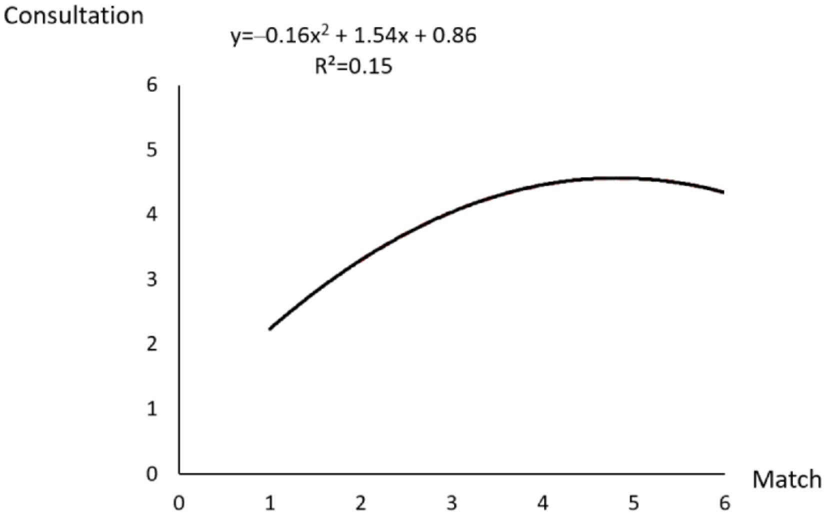
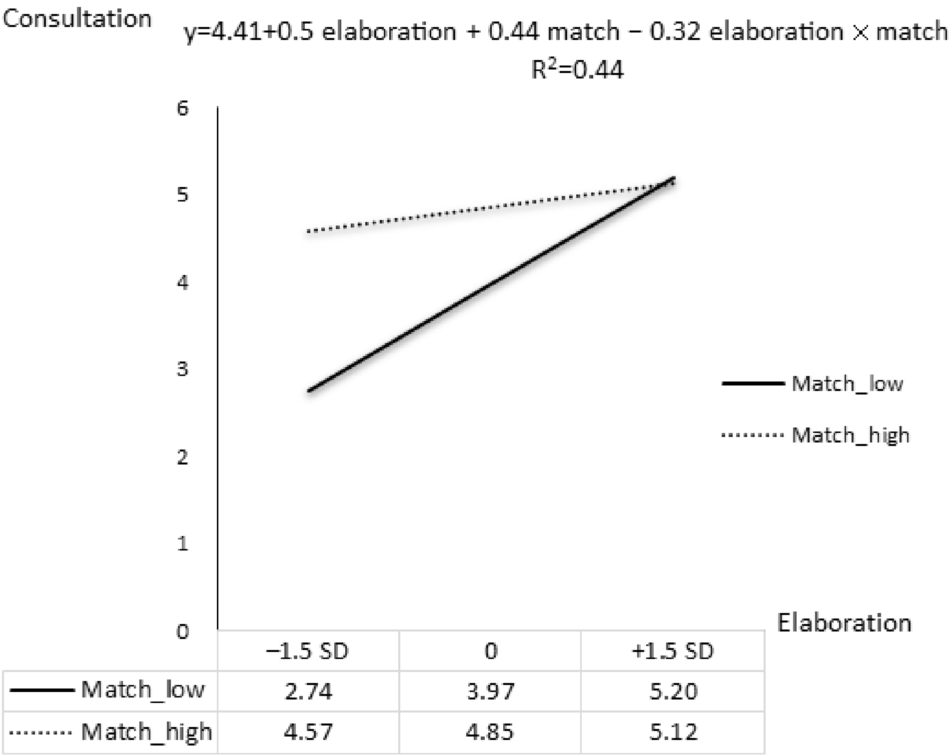


**Figure 2.** Nonlinear graph of the match variable.



To better understand the interaction effect between elaboration and match, we used the standardized latent means of elaboration and match calculated by PLS to analyze the regression of consultation for representative groups. As suggested, the low-and the high-match group were chosen at low (–1 SD from the mean) and high (1 SD from the mean) values of match, respectively [43,44]. To observe the crossover interaction, the consultation scores for the low- and high-match groups were calculated at a low level (–1.5 SD) and a high level (1.5 SD) of elaboration, respectively [44]. Next, the predicted values for each group were produced by multiplying the respective unstandardized regression coefficients for each variable at an appropriate value (eg, high match=1, high elaboration=1.5). Figure 3 shows the plot of interaction between match and elaboration. The solid line is for the low-match group (at a value of –1), while the dotted line is for the high-match group (at a value of 1). The result indicates that elaboration had a positive effect on consultation for both the low- and the high-match group. However, the slopes show that when the match was low, the effect of elaboration on consultation was stronger than that of a high match. Furthermore, the crossover interaction shows that when elaboration was low, the high-match group had a higher consultation score than that of the low-match group. On the contrary, when elaboration was high, the low-match group had a higher consultation extent than that of the high-match group.

**Figure 3.** Interaction effect between match and elaboration on consultation.



| | –1.5 SD | 0 | +1.5 SD |
|---|---|---|---|
| Match_low | 2.74 | 3.97 | 5.20 |
| Match_high | 4.57 | 4.85 | 5.12 |

## Discussion

### Principal Findings

#### *Role of Education in Health Information Seeking*

It has been reported that online health information may potentially benefit individuals by making them better informed, resulting in more effective health outcomes; on the contrary, misinformed health information may result in inappropriate use of medical resources [15]. In addition, studies have indicated that individuals with lower education levels are less likely to access websites for health information and show unsuitable utilization behaviors, while people with university degrees more frequently access online health information using complex and expanded information search strategies [10,25,26]. Therefore, the population without university degrees has been regarded as an important target group to examine their online health information navigation behaviors [22]. Accordingly, the results of this study may provide expanded views on the online health information search behaviors of those with low educational backgrounds.

#### *Positive Influences of Health Information Searching on Consultations*

As can be seen from the results presented in Table 1, the correlations between information search strategies and consultation showed that both match strategy and elaboration strategy have positive influences on the usage of online health information to consult others. That is, no matter what search strategy the patients used to gather online health information, they were willing to further discuss the information with medical experts or others. Despite an advanced search strategy, such as an analytic approach, being considered an important factor connected with accurate search results, it was emphasized that simple strategies, such as the browsing approach, which may be efficient and successful, need not necessarily be rejected [23]. As was expected, patients with more exposure to health information through information communication technology (both advanced and simple approaches) were more likely to perform healthier behaviors, suggesting a potential way for health care professionals to encourage their patients to access online health information and communicate health information with them through digital media [10].

### *Curvilinear Relationship of Match With Consultation*

Curvilinear analysis of the match strategy indicated that it was positively linked to consultation willingness, but the correlation became weaker and even changed direction as the match strategy reached a high level. That is, accessing online health information through the match strategy is necessary and helpful for consulting health care professionals about the retrieved information, but too much use of this simple approach may disadvantage consultation behaviors. Similarly, it was reported that health information seekers without medical expertise are more likely to use search engines to perform a simple search; although it would be useful to engage them in the information discovery process, it also may become a barrier to further obtaining the most suitable solution [20].

Based on the theory of planned behavior (TPB), it was indicated that abundant information may overload information seekers and result in their psychological ill-being (eg, depression and anxiety), which may discontinue their intention to use the online health information [45]. Accordingly, it may explain why the match strategy has a positive influence on consultation behavior, implying that gathering relevant information from a few resources may support information seekers' continuous use of online health information. Nevertheless, an overwhelming amount of information retrieved by the match strategy without the skills of evaluating and integrating such information may discourage its continuous use. To summarize prior research, there are interesting findings on health information seekers' health information–seeking behaviors and responses to the gathered information [5,20,31]. Simple lookup search strategies may have both advantages and disadvantages for individuals' health information–seeking behaviors [5,20,31]. Furthermore, it was demonstrated that multiple health information sources through an instant search approach can lead to information overload and result in information avoidance, suggesting the need for training on advanced health information–seeking skills to manage and integrate diverse information sources [46].

### *The Elaboration Strategy Is Essential to Desired Health-Seeking Behavior*

As laypeople do not have medical expertise, they tend to adopt basic search strategies to look up online health information for retrieving facts and answering health questions [20,47]. However, the correlation analyses in this study showed that the elaboration strategy has more positive influences on consultation than the match strategy. In addition, interaction effect analysis indicated the important role of the elaboration strategy in reinforcing patients' willingness to further consult medical experts or others with the online health information they have found, especially patients with a tendency to adopt a low-match strategy. In conclusion, the elaboration strategy may be a better choice than the match strategy through which to encourage patients to gather and integrate numerous types of health information and use such information appropriately. To further understand and interpret health information, health information seekers have to adopt advanced search strategies to scan and justify the search results [47]. As suggested, patients and their relatives were encouraged to conduct more advanced search strategies to recognize credible and appropriate health information sources [32].

### *The Importance of eHealth Literacy for Advancing Health-Seeking Behavior*

In Taiwan, an investigation on health information–seeking behaviors showed that internet users with high educational levels (university and above) are more likely to use the internet for health information searching. Regarding the effects of health information searching, a majority of the respondents used such health information to ask physicians questions and to make decisions on disease treatment and whether to consult a physician [4]. Therefore, online health information seeking can be regarded as a channel through which health care professionals can enhance patient-physician relationships and help patients by recommending credible health information sources.

In conclusion, there is a need to investigate how to stimulate internet users with low educational levels (without a university education) to use health information to consult health care professionals and to have positive effects on their treatment decisions and health outcomes. Based on the results of this study, health care professionals may better know how medicine-related information search strategies (ie, match and elaboration strategies) can benefit patients with low educational levels when turning to the internet for making health decisions [26]. In sum, this subpopulation (those with less education) may benefit from online information only when they have access to alternative health information sources, such as health care providers [30].

When compared with the low-level-eHealth-literate group, high-level-eHealth literate individuals who have a good ability to seek, locate, evaluate, and apply online health information were recognized as more frequent health information seekers and were better at using effective online health information search strategies to address their health concerns [5]. As suggested, improving eHealth literacy may promote individuals' use of effective online information-seeking strategies and identify high-quality health information sources. In the case of this study, for patients in both the low- and the high-match group (in particular, those with a low tendency to adopt the match strategy), developing their eHealth literacy may encourage their intent to use the elaboration strategy and consult health care professionals.

## Limitations

Several limitations of this study should be noted. First, this study targeted laypeople without a university education in order to examine their online health information search behaviors rather than other populations with a university degree or higher educational background. That is, the results of this study should be cautiously interpreted and inferences should be made with care. The second limitation is the sampling method used in this study. The participants included in this study were purposefully recruited from 1 university-affiliated teaching hospital rather than from other clinical settings, such as small hospitals or private clinics. Therefore, the generalizability of the study results is limited to other clinic settings and regions in Taiwan. Third, instead of objective data, such as log files, the data of this study were collected from patients' subjective opinions and attitudes. Thus, the self-reported bias should be considered. Finally, a few predicting factors, including age, sex, search frequency, and search strategies, were explored in this study and recruited in the regression analysis model. Although the results of PLS path analysis indicated that a high proportion of variances was explained by the predictors, there is still a need to further consider other predictors or confounding factors, such as severity of illness and accessibility of medical resources, which may influence how patients use online health information.

## Conclusion

Although there are challenges for laypeople, who are not medical experts, and who do not have a university degree to properly access and evaluate the credibility and accuracy of health information retrieved from the internet [10,22], understanding their online health information search strategies and use of such information may help health care professionals better know how to lead their patients to appropriately search for and communicate about online health information with medical experts. Certainly, the internet is an essential tool through which patients may approach the low-cost wealth of health information; however, it is an additional source of health information, which should not necessarily replace traditional health information offered by health care professionals [9].

Based on the findings of this study, we provided practical suggestions in several aspects. As suggested, the public population and patients were encouraged to gather health information from multiple sources, including medical experts' advice, as well as alternative opinions from the internet [7,25]. It has been indicated that patients use online medical information to integrate with advice from friends, family, and physicians in order to confidently make their medical decisions [12]. According to the results of this study, patients without a university degree should be supported to obtain more exposure to online health information through both complex and simple search approaches, which in turn may induce them to consult medical experts about such information. In addition, it was suggested that health care providers should recognize that their patients are using the internet as a medical information source, and should be prepared to help patients to carefully identify the quality of online health information and appropriately use such information [48]. That is, medical professionals must be aware that they are eligible to direct patients' health information search behaviors and empower them to engage in an informed and active way in their own medical decision-making process. Finally, we recommend that health care providers offer high-quality information on well-designed medical websites. To assist patients in adopting simple searches and becoming advanced explorers, there is a need to provide better information tools and quality content for them to surf the internet full of rich information and many pitfalls [47]. While conducting a heuristic search, patients sometimes reject credible websites with high-quality content due to poor visual appeal and unclear interface design [12]. In other words, well-designed websites built by medical professionals containing a clear interface and quality health information can draw the attention of patients and lead them to access trustworthy information while looking up health information on the internet.

Meanwhile, the results of the interaction and curvilinear analyses suggested that the elaboration strategy is a more recommendable approach than the match strategy through which patients are more likely to use online health information to consult with their doctors or others about their health-related problems. To stimulate patients' online health information search strategies in more advanced ways, it has been suggested that advancing patients' eHealth literacy (ie, ability to search, locate, evaluate, integrate, and apply electronic health information) may support them to conduct appropriate information search strategies, justify reliable and useful information, and use such information in an effective manner [5,45].

In summary, this study acknowledges how patients without a university degree search for health information over the internet, how they share the information with doctors and others, and how to guide them to accurately use the information sources. As patients have better access to additional medical advice over the internet and can discuss such information with health care professionals, they are expected to be more involved in appropriate health information and engaged in their medical decision making.

## Acknowledgments

This work was, in part, financially supported by the Institute for Research Excellence in Learning Sciences of the National Taiwan Normal University (NTNU) through the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. In addition, the authors are thankful for the research funding offered by the Ministry of Science and Technology (MOST), Taiwan (Grant MOST 110-2628-H-002 -004 -MY2 and MOST 108-2511-H-003 -004 -MY3). We also thank all those involved in this research, including the patients who replied to the questionnaires and the research assistants who helped with collecting the questionnaires.

## Authors' Contributions

Y-LC contributed to the study design, developed the instruments, analyzed the research data, and also drafted the main text of this paper. C-CT contributed to constructing the research model. He also gave opinions and interpretations to explain the results of the statistical analysis. J-CL developed the instruments and contributed to the study design. In addition, he assisted in delivering and collecting the questionnaires and provided opinions on the results of the statistical analysis.

## Conflicts of Interest

None declared.

---

Multimedia Appendix 1
Details about the questionnaire and its validation.
[DOCX File , 17 KB - jmir_v24i9e29609_app1.docx ]

---

Multimedia Appendix 2
Cover letter describing informed consent.
[DOCX File , 13 KB - jmir_v24i9e29609_app2.docx ]

---

## References

1. Prestin A, Vieux SN, Chou WS. Is online health activity alive and well or flatlining? Findings from 10 years of the Health Information National Trends Survey. J Health Commun 2015 Jul;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
2. Fox S, Duggan M. Health Online 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2022-03-23]
3. TWNIC. A Survey on Broadband Internet Usage in Taiwan: A Summary Report. URL: https://www.twnic.tw/download/200307/20150901d.pdf [accessed 2022-03-23]
4. Koo M, Lu M, Lin S. Predictors of internet use for health information among male and female internet users: findings from the 2009 Taiwan National Health Interview Survey. Int J Med Inform 2016 Oct;94:155-163. [doi: 10.1016/j.ijmedinf.2016.07.011] [Medline: 27573323]
5. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Comput Hum Behav 2017 Apr;69:256-267. [doi: 10.1016/j.chb.2016.12.032]
6. Bratucu R, Gheorghe IR, Purcarea RM, Gheorghe CM, Popa Velea O, Purcarea VL. Cause and effect: the linkage between the health information seeking behavior and the online environment; a review. J Med Life 2014 Sep 15;7(3):310-316 [FREE Full text] [Medline: 25408746]
7. Chen Y, Li C, Liang J, Tsai C. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. J Med Internet Res 2018 Feb 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]
8. McMullan M. Patients using the internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
9. Mano R. Online health information, situational effects and health changes among e-patients in Israel: a 'push/pull' perspective. Health Expect 2015 Dec 17;18(6):2489-2500 [FREE Full text] [doi: 10.1111/hex.12218] [Medline: 24938381]
10. Shen C, Wang MP, Wan A, Viswanath K, Chan SSC, Lam TH. Health information exposure from information and communication technologies and its associations with health behaviors: population-based survey. Prev Med 2018 Aug;113:140-146. [doi: 10.1016/j.ypmed.2018.05.018] [Medline: 29787775]
11. Xie B, Wang M, Feldman R, Zhou L. Internet use frequency and patient-centered care: measuring patient preferences for participation using the health information wants questionnaire. J Med Internet Res 2013 Jul 01;15(7):e132 [FREE Full text] [doi: 10.2196/jmir.2615] [Medline: 23816979]
12. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
13. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
14. Medlock S, Eslami S, Askari M, Sent D, de Rooij SE, Abu-Hanna A. The consequences of seniors seeking health information using the internet and other sources. Stud Health Technol Inform 2013;192:457-460. [Medline: 23920596]
15. Chen Y, Chen L, Kao Y, Chu T, Huang T, Ko W. The over-optimistic portrayal of life-supporting treatments in newspapers and on the internet: a cross-sectional study using extra-corporeal membrane oxygenation as an example. BMC Med Ethics 2014 Aug 01;15:59 [FREE Full text] [doi: 10.1186/1472-6939-15-59] [Medline: 25081491]
16. Ahmad F, Hudak PL, Bercovitz K, Hollenberg E, Levinson W. Are physicians ready for patients with internet-based health information? J Med Internet Res 2006 Sep 29;8(3):e22 [FREE Full text] [doi: 10.2196/jmir.8.3.e22] [Medline: 17032638]
17. Wald HS, Dube CE, Anthony DC. Untangling the web: the impact of internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]
18. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]

19. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
20. Pang PC, Verspoor K, Chang S, Pearce J. Conceptualising health information seeking behaviours and exploratory search: result of a qualitative study. Health Technol 2015 Feb 10;5(1):45-55. [doi: 10.1007/s12553-015-0096-0]
21. Kammerer Y, Bråten I, Gerjets P, Strømsø HI. The role of internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. Comput Hum Behav 2013 May;29(3):1193-1203. [doi: 10.1016/j.chb.2012.10.012] [Medline: 25904163]
22. Kammerer Y, Amann DG, Gerjets P. When adults without university education search the internet for health information: the roles of internet-specific epistemic beliefs and a source evaluation intervention. Comput Hum Behav 2015 Jul;48:297-309. [doi: 10.1016/j.chb.2015.01.045]
23. Sharit J, Taha J, Berkowsky RW, Profita H, Czaja SJ. Online information search performance and search strategies in a health problem-solving scenario. J Cogn Eng Decis Mak 2015 May 21;9(3):211-228 [FREE Full text] [doi: 10.1177/1555343415583747] [Medline: 29056885]
24. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
25. AlGhamdi KM, Moussa NA. Internet use by the public to search for health-related information. Int J Med Inform 2012 Jun;81(6):363-373. [doi: 10.1016/j.ijmedinf.2011.12.004] [Medline: 22217800]
26. Perez SL, Kravitz RL, Bell RA, Chan MS, Paterniti DA. Characterizing internet health information seeking strategies by socioeconomic status: a mixed methods approach. BMC Med Inform Decis Mak 2016 Aug 09;16(1):107 [FREE Full text] [doi: 10.1186/s12911-016-0344-x] [Medline: 27506607]
27. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013 Jan 30;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
28. Thapa DK, Visentin DC, Kornhaber R, West S, Cleary M. The influence of online health information on health decisions: a systematic review. Patient Educ Couns 2021 Apr;104(4):770-784. [doi: 10.1016/j.pec.2020.11.016] [Medline: 33358253]
29. Zimmerman MS, Shaw G. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020 Sep 12;37(3):173-191. [doi: 10.1111/hir.12287] [Medline: 32052549]
30. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
31. Wang HY, Liang JC, Tsai CC. The difference in the online medical information searching behaviors of hospital patients and their relatives versus the general public. J Educ Technol Soc 2014;17(3):280-290 [FREE Full text]
32. Larner AJ. Searching the internet for medical information: frequency over time and by age and gender in an outpatient population in the UK. J Telemed Telecare 2006 Jun 24;12(4):186-188. [doi: 10.1258/135763306777488816] [Medline: 16774699]
33. Alhuwail D, Abdulsalam Y. Assessing electronic health literacy in the state of Kuwait: survey of internet users from an Arab state. J Med Internet Res 2019 May 24;21(5):e11174 [FREE Full text] [doi: 10.2196/11174] [Medline: 31127723]
34. Wu YT, Tsai CC. Developing an Information Commitment Survey for assessing students' web information searching strategies and evaluative standards for web materials. Educ Technol Soc 2007;10(2):120-132 [FREE Full text]
35. Tsai C. Information commitments in web-based learning environments. Innov Educ Teach Int 2011 Jun 20;41(1):105-112. [doi: 10.1080/1470329032000172748a]
36. Hair Jr JF, Matthews LM, Matthews RL, Sarstedt M. PLS-SEM or CB-SEM: updated guidelines on which method to use. IJMDA 2017;1(2):107. [doi: 10.1504/ijmda.2017.10008574]
37. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39. [doi: 10.2307/3151312]
38. Hair JJ, Black W, Babin B, Anderson R, Tatham R. Multivariate Data Analysis. 6th Ed. NJ: Prentice Hall; 2006.
39. Human Subjects and Research Act. Laws and Regulations Database of the Republic of China (Taiwan). 2019 Jan 02. URL: https://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=L0020176
40. National Development Council. Department of Health, Executive Yuan Notice is hereby given, for the promulgation of "Exempt Review Categories for Human Research" (promulgation becomes effective from 5th, July 2012). The Executive Yuan Gazette Online. 2012 Jul 05. URL: https://gazette.nat.gov.tw/egFront/e_detail.do?metaid=54295
41. Hair Jr JF, Hult GTM, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2nd Ed. CA: SAGE Publications; 2017.
42. Basco R, Hair JF, Ringle CM, Sarstedt M. Advancing family business research through modeling nonlinear relationships: comparing PLS-SEM and multiple regression. J Fam Bus Strategy 2021 Oct:100457. [doi: 10.1016/j.jfbs.2021.100457]
43. Frazier PA, Tix AP, Barron KE. Testing moderator and mediator effects in counseling psychology research. J Couns Psychol 2004;51(1):115-134. [doi: 10.1037/0022-0167.51.1.115]
44. Aiken LS, West SG. Multiple Regression: Testing and Interpreting Interactions. Thousand Oaks, CA: Sage Publications; Sep 03, 1991:167-168.

45. Swar B, Hameed T, Reychav I. Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search. Comput Hum Behav 2017 May;70:416-425. [doi: 10.1016/j.chb.2016.12.068]
46. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara S. From information seeking to information avoidance: understanding the health information behavior during a global health crisis. Inf Process Manag 2021 Mar;58(2):102440 [FREE Full text] [doi: 10.1016/j.ipm.2020.102440] [Medline: 33281273]
47. Marchionini G. Exploratory search. Commun ACM 2006 Apr 01;49(4):41-46. [doi: 10.1145/1121949.1121979]
48. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]

## Abbreviations

**AVE:** average variance extracted
**CR:** composite reliability
**ICS:** Information Commitment Survey
**OHIU:** Online Health Information Utilization
**PLS-SEM:** partial least squares-structural equation modeling

*Edited by G Eysenbach; submitted 15.04.21; peer-reviewed by J Taylor, W Pian; comments to author 28.06.21; revised version received 25.08.21; accepted 06.03.22; published 02.09.22.*

*Please cite as:*
*Chiu YL, Tsai CC, Liang JC*
*Laypeople's Online Health Information Search Strategies and Use for Health-Related Problems: Cross-sectional Survey*
*J Med Internet Res 2022;24(9):e29609*
*URL: https://www.jmir.org/2022/9/e29609*
*doi:10.2196/29609*
*PMID:36053561*

Original Paper

# Next-Generation Capabilities in Trusted Research Environments: Interview Study

Sanaz Kavianpour[1], PhD; James Sutherland[2], PhD; Esma Mansouri-Benssassi[2], PhD; Natalie Coull[1], PhD; Emily Jefferson[2], PhD

[1]School of Design and Informatics, Abertay University, Dundee, United Kingdom
[2]Health Informatics Centre, University of Dundee, Dundee, United Kingdom

**Corresponding Author:**
James Sutherland, PhD
Health Informatics Centre
University of Dundee
Dundee, DD1 4HN
United Kingdom
Phone: 1 5153054219
Email: j.a.sutherland@dundee.ac.uk

## Abstract

**Background:** A Trusted Research Environment (TRE; also known as a Safe Haven) is an environment supported by trained staff and agreed processes (principles and standards), providing access to data for research while protecting patient confidentiality. Accessing sensitive data without compromising the privacy and security of the data is a complex process.

**Objective:** This paper presents the security measures, administrative procedures, and technical approaches adopted by TREs.

**Methods:** We contacted 73 TRE operators, 22 (30%) of whom, in the United Kingdom and internationally, agreed to be interviewed remotely under a nondisclosure agreement and to complete a questionnaire about their TRE.

**Results:** We observed many similar processes and standards that TREs follow to adhere to the Seven Safes principles. The security processes and TRE capabilities for supporting observational studies using classical statistical methods were mature, and the requirements were well understood. However, we identified limitations in the security measures and capabilities of TREs to support “next-generation” requirements such as wide ranges of data types, ability to develop artificial intelligence algorithms and software within the environment, handling of big data, and timely import and export of data.

**Conclusions:** We found a lack of software or other automation tools to support the community and limited knowledge of how to meet the next-generation requirements from the research community. Disclosure control for exporting artificial intelligence algorithms and software was found to be particularly challenging, and there is a clear need for additional controls to support this capability within TREs.

***(J Med Internet Res 2022;24(9):e33720)*** doi:10.2196/33720

**KEYWORDS**

data safe haven; health data analysis; trusted research environment; TRE

## Introduction

A Trusted Research Environment (TRE), sometimes also known as a Data Safe Haven, Virtual Research Data Center, or Virtual Data Enclave, is a secure environment designed for approved and named researchers to access sensitive data, where access to specific data sets is provided to approved research projects. To protect the confidentiality and privacy of the data, TRE providers and researchers using the environments generally follow a set of TRE principles. Such principles have developed over time, for example, the Scottish Safe Haven Charter and the Health Data Research Alliance Trusted Research Environment Green Paper [1-3].

The objective of a TRE is to provide safe and trustworthy access to data for research. Controls are generally applied to both the import and export of data to protect the privacy of data subjects and the integrity of the environment itself. For example, a researcher may be given access to a data table listing the age, medication taken, and hospital admission information on each of 5000 people. They will be allowed to identify and publish aggregate statistical observations that people aged >60 years who are taking a particular drug were x% less likely to be admitted to the hospital with heart problems, but the details of each individual will never leave the TRE.

Within the secure environment, researchers can analyze these data using a set of advanced analytics tools, for example, R and SPSS. Some TREs also offer the researcher the capability to program within the environment, support the development of new artificial intelligence (AI), and apply natural language processing for the analysis of unstructured text.

Many TREs have been developed to host health data. For example, in the United Kingdom, several TREs are established to host health records from the National Health Service (NHS), the publicly funded health care system of the United Kingdom [4-8]. A similar model has since been adopted to provide secure access to many other non–health-related data sets [9-11]. Many TREs now regularly host both health-related and non–health-related data. Providing researchers access to sensitive data sources without compromising the privacy and security of the data is a complex process. Historically, TREs have mainly supported observational studies on text-based structured data using standard statistical packages. There is a growing requirement from the research community (academic and industrial) for TREs to provide additional capability beyond simple support for observational data statistical analysis, but without compromising the security or privacy of the data. In this paper, we term these requirements as next-generation TRE capabilities. These include the following:

1. Support for big, nonstructured data (such as genomic and imaging data, which can be several terabytes in size)
2. Ability to parallelize computational jobs to either a high-performance computing cluster or a graphics processing unit (GPU) farm
3. Support for software development within the TRE
4. Freedom to install software packages of researcher choice
5. Ability to export software and AI algorithms from the environment
6. Ability to connect to certain internet locations, for example, code repositories (GitHub)

This study aims to understand the state of the art in supporting next-generation TRE capabilities; existing technical security measures that have been adopted and how widely; and limitations in existing controls and processes, where active studies are required to develop novel methods.

## Methods

### Participant Recruitment

A total of 73 individuals from 42 different organizations (n=22, 52% in the United Kingdom and n=20, 48% overseas) were invited to participate, but response rates were substantially higher in the UK cohort (20/73, 27% responses vs 7/73, 10% responses), perhaps owing to greater familiarity. A few respondents either declined or were unable to participate for practical reasons, and 3% (2/73) of them felt that their situation was already covered by other participants from related institutions, so they were not counted separately.

The findings are based on interviews conducted with 84% (16/19) of TRE providers in the United Kingdom and 16% (3/19) of TRE providers across Canada, Australia, and Europe. Each interview took approximately 2 hours, using a set of questions that were designed to cover TRE controls and next-generation capabilities (Multimedia Appendix 1).

Building upon the Five Safes model [12], the recent HDR UK Green Paper [2] describes seven “Safes”: safe people, safe projects, safe setting, safe computing (an extension of safe setting), safe data, safe outputs, and safe return (extending the TRE definition). This study focuses on a subset of the controls that will support next-generation TRE capabilities (data, outputs, settings, computing, and people). Textbox 1 provides a summary of the topics that were discussed with the participants during the interview, under each of our subset of the 7 Safes.

The interview participants were recruited mainly from the technical TRE infrastructure teams. A nondisclosure agreement was signed by the project parties, to provide assurance for any participant who chooses to disclose information of confidential or proprietary nature. We have anonymized the responses in this paper and grouped them under different Safes. There were some differences in some controls and measures from TREs from different countries, but we could not directly highlight these as they may have affected the anonymity of participants.

Where relevant, we also present our analysis of the TRE limitations identified and recommendations that we believe can help improve TREs.

**Textbox 1.** Safes and discussion points.

**Safe data**

- Tools and techniques used to manage and reduce the potential risk of reidentification by applying disclosure control to all data imported to the Trusted Research Environment (TRE)

**Safe outputs**

- Types of data that can be exported from the TRE
- Future plans to enable export of additional types of data
- The process used for checking disclosure control on data to be exported, including frequency and restrictions
- Software and any manual checks used for disclosure control

**Safe settings**

- Standard build of the TRE (including computing power, operating system, and software)
- Maximum computing power offered to TRE "power" users
- Security measures used to mitigate the risk of unauthorized access, data loss, and misuse
- Rules regarding the import of data or code (including libraries) into the environment
- Support for federated queries of data from external sources

**Safe computing**

- Use of private (dedicated hardware) or public cloud (paying for capacity from a service such as Amazon Web Services or Azure)

**Safe people**

- The controls put on the people who use the TRE
- Whether access to the environment must be via a recognized "trusted" organization

## Ethics Approval

The work presented in this paper was considered by Abertay University Ethics Committee on May 28, 2020, and granted full approval, with reference number EMS3012.

# Results

## Safe Data

### *Overview*

The principles of safe data relate to the data allowed to be imported into the TRE. Good practice indicates that such data should be of high quality and pseudonymized or anonymized [2]. Researchers accessing the TRE should only be able to access the data necessary for their research project; the work of protecting data begins by applying disclosure control to control and assess data provided to the researchers within the TRE, before a second pass at the export stage regulates the data those researchers are allowed to disclose publicly within their research output.

Many TREs provide a service to link and anonymize data from different data sources. Such research projects often require data governance approval before researchers can access the data. TREs often require researchers to sign data user agreements, which reinforce the rules and consequences of violations. During the interviews, participants were asked to discuss the existing tools used to support safe data and their views on future solutions. These are discussed in the following sections.

### *Breaches or Near-Miss Incidents*

Each participating TRE has processes in place for incident response, including reporting to the appropriate authorities when required. None of them reported any actual reportable incidents. In total, 9% (2/22) of the participants acknowledged that there can be lapses of procedure (these are different from data breaches and do not need to be reported to the Information Commissioner's Office), where a researcher may request to export data that are not permitted, perhaps owing to researcher error. In these cases, the incident will be addressed with a process for review, formal warning, and retraining, if necessary. One participant described an incident in which a field that normally contains facility or hospital names was provided to a researcher, and some records included private addresses in this field. Then, these data were retracted and resupplied with a placeholder for private addresses. In this example, the data were not released outside the TRE, but researchers saw the potentially identifiable data within the TRE environment.

### *Tools and Techniques Used to Manage the Risks of Reidentification*

#### Overview

Most participants (16/22, 73%) do not rely on special purpose disclosure control tools, but on their analysts' knowledge and communication about the purpose of the project and the nature of the data. Ultimately, TREs exist with the aim of providing "safe data" sufficient for the project's needs. Generally, patient IDs are replaced with either TRE-specific or project-specific identifiers; where possible, other identifying information is redacted or reduced in resolution, for example, replacing a date of birth with a year or month of birth. The combination of data across projects is not allowed and is made impractical through the use of project-specific pseudonyms in place of original identifiers.

A total of 18% (4/22) of the participants explicitly measure disclosure risks at the import stage for each project; however, all participants (22/22, 100%) acknowledged that disclosure risk is an important factor when evaluating a new project application. As part of this process, a TRE has all data scrutinized by an external organization, using the k-value (the minimum number of individuals sharing any combination of identifying characteristics) to quantify the reidentification risk for each data set [13]. For example, a data set classifying patients by age and gender, with a minimum of 4 patients in each category, has k=4. Grouping patients into larger age bands will increase the k-value, thus reducing risk at the expense of reduced data resolution. TREs apply a similar measure as part of the export process, usually requiring a minimum of 5 to 10 individuals in any output grouping to ensure that no individual can be reidentified, as detailed in section 3.2.

A total of 18% (4/22) of the participants reported using tools to support checks during the import and export processes. Each of them used a different tool, as follows:

1. sdcMicro (as discussed in section 2.6) is a free R-based open-source package that assesses the risk of a data set containing identifiable data using various risk estimation methods [14].
2. Privitar's Data Privacy Platform product applies user-defined policies for filtering and transforming data, including adding random "noise" to numerical values in configurable ways to reduce identifiability [15].
3. Custodix Anonymisation Tool Services deidentification platform [16] provides both assessment and deidentifying transformation of various types of data (eg, comma separated variables and Digital Imaging and Communications in Medicine).
4. Privacy Analytics Risk Assessment Tool [17] is used by a TRE to assess risk, redact, and deidentify source data and to assess export requests.

### Recommendations

There is a need for affordable tools that can be used in TREs to support deidentification of data and assess the risk of reidentification; in particular, building on the existing use of the free sdcMicro toolkit should be considered. Data identifiability is not binary [18], and data can be identified indirectly by combining attributes, which is known as a triangulation attack [19].

## Safe Outputs

### Overview

Participants were asked which data types can be exported from their TRE and what controls are placed on the export of data. These checks are more extensive than those for imported data, adding checks for deliberate attempts to hide data, for example, as white text in the document or embedded in Stata code and generic checks on the actual data going further than the input checks—usually an absolute prohibition on data regarding any individual ("row-level").

Generally, the researcher explicitly requests the export of specific files, which are then reviewed by TRE staff, and in some cases, other relevant external parties (eg, data owners) before the permission to export is granted. All the participants have instructions that document the manual checks required. The team determines the extent of checks required at the data or project level based on the sensitivity of the data. For example, openly available, public data sets will not typically require an independent referee, but clinical data may require more consideration. To obtain highly sensitive data, multiple members of the review team may be required to check each file.

The checks needed varied between TREs and sometimes between projects, depending on the nature of the data, including how sensitive it is and whether it was consented or not, with release criteria being agreed upon between data owners and TRE operators, sometimes consulting with research teams for specific situations and then enforced by TRE staff. Many TREs have developed a rule-based framework to categorize projects and data into specific types. For example, open public data can be exported with minimal checks, whereas for clinical data, only aggregate-level summary data can be exported. Others used a simpler *one-size-fits-all* approach.

### Export of Individual-Level Data

Generally, export of row or individual-level data is permitted only for projects where the data are already openly available in the public domain or where specific consent had been provided by the study participants to collect and share the data. In the latter case, respondents indicated that the data controller was most likely to be the principal investigator of the project, and they will typically use the TRE to securely manage access to their data by different researchers involved in the project, but will choose not to place restrictions on data export. Otherwise, only aggregate statistics can be exported.

### Export of Aggregate-Level Statistical Analysis

Most TREs only allow export of aggregate-level statistical analysis. For example, in clinical data projects where the data controller is an NHS board or trust, researchers are not permitted to export any data related to specific individuals (even if pseudonymized).

All participants (22/22, 100%) indicated that their TREs allow the export of aggregate-level data as graphs or tables, with a minimum number of data points in any table cell or graphical output to reduce the probability of reidentification of data from small sample sizes ("small cell risk"). Of the 22 participants, 17 (77%) participants have set policies in place: 32% (7/22) of the participants reported using a minimum of 5 individuals in a cell, with 9% (2/22) of the participants using 10 and 36% (8/22) varying the limit depending on the context of the study and the nature of the underlying data.

All participants (22/22, 100%) acknowledged that there was a potential risk of reidentification if sufficient data points are exported from the TRE, known as jigsaw identification or triangulation [20]. Although there is a clear need for researchers to export aggregate-level statistical analysis, the mitigations used for these risks vary across different data sets [3]. Software tools can be used to estimate the probability of reidentification, as discussed in section 2.2.

Introducing new types of data as export options brings new risks, particularly AI models and software, where statistical analysts will be unfamiliar with the nature of such files and manual inspection is ineffective, difficult, and time-consuming, thus introducing new security risks. Data can be intentionally concealed within such files and recovered by a third party from an exported model created innocently through means such as membership inference attacks [21] or, with some internal knowledge or collusion with a researcher, inversion attacks to deduce additional information about the training data [22].

In the following section, we discuss the different data types that can be exported from the participants' TREs and how this is managed.

## Export of AI Algorithms, Software, and Scripts

### Overview

Participants were asked about policies regarding the export of software and AI models developed within their TREs. Of the 22 participants, 5 (23%) allow export of AI models, 8 (36%) specifically prohibit this, and 5 (23%) are prepared to consider it in the future.

Some participants also plan to support the export of R and Stata scripts in the future, if they have established a suitable process for reviewing.

A total of 74% (14/19) of the TREs permit the export of software source code developed within the TRE. None of them (0/19, 0%) have been asked to allow compiled executables, 11% (2/19) are prepared to consider this with safeguards. However, it should also be entirely avoidable by developing the source code outside the TRE and then deploying it into the TRE for testing [23].

### Limitations

Many participants (9/22, 41%) indicated that checking algorithms, software, and scripts is very challenging because a malicious individual can "hide" individual-level data within the files. For example, the weights of an AI algorithm are a set of numbers and sensitive data can be embedded in them. This is very difficult to detect, particularly if a malicious user disguises the data. It is also possible to include individual-level data inadvertently, for example, if the AI algorithm is overtrained, the weights correspond to the data underneath, or if an R script incorporates the underpinning data. Checking a substantial software project manually is unrealistic.

### Recommendations

Developing AI models in TREs without compromising patient privacy requires tools such as those proposed by Nicolae et al [24] and Liu et al [25] to quantify their risk and vulnerability to attacks (eg, membership inference attacks [26,27], deanonymization attacks [28], reconstruction attacks [29], model extraction attacks [30,31], and model inversion attacks [23,32]) and consider integrating privacy mechanisms in the model development to counter these attacks [33]. Best practice guidelines can also help users to design robust and safe algorithms, including through auditable and explainable AI [34]. Software tools to check for nonmalicious export by comparing individual data within the TRE with those in the export files is a possibility, but such tools are not currently routinely used by any of the TREs. Barriers to their use by TREs include the attack-specific nature of such tools and their high price. For software development (as opposed to AI models, where the training data are essential input for the end product), exporting the software from the TRE can be avoided entirely by developing it outside.

## Automation of Data Export Checks

### Overview

Although software can theoretically be used to facilitate the data export process, none of the participants (0/22, 0%) believed that software can currently replace the role of humans for checking export files. A participant questioned whether it would ever be feasible to fully automate all aspects of the process, largely owing to concerns about trusting the software to perform all the necessary checks without human oversight. Some participants (8/22, 36%) felt that the software available is currently not sufficiently mature to manage all the risks and humans are better at the task; however, they indicated that they would be willing to incorporate software of this nature into the process in the future as the technology evolves.

Of the 22 participants, 2 (9%) participants reported using automated tools for export checks. As noted in section 2.2, one participant used the proprietary Privacy Analytics Risk Assessment Tool product and another participant used a simple in-house tool to detect the project-specific identifiers they use, but the main disclosure checks are manual. Currently, the use of automated tools seems to be more prevalent for import checks than for export checks.

### Limitations

Manual checks are time-consuming and error-prone, with a risk of missing concealed data (steganographic, white-on-text, and undo buffers) and delays in data release. Although the participants acknowledged that the current data center process can be enhanced with automated tools, there are significant concerns with relying solely on technology to check export requests, based on the potential ramifications for any unapproved data to accidentally leave the environment and the challenges with checking algorithms. Proprietary tools are expensive, and TREs try to keep the costs low for academic research.

### Recommendations

A hybrid model with automated checks can facilitate and accelerate export and reduce the risk of reidentification, checking more thoroughly for inadvertent and malicious inclusion of data. The tools mentioned in section 2.2 may also be useful in this role. Best practice guidance regarding methods to reduce the opportunities for malicious data exfiltration can also help. Although governance (section 6) can help to ensure that researchers are trustworthy, malicious attempts to hide data should be considered, for example, in the event of stolen researcher credentials.

### *Frequency of Data Exports*

#### Overview

All participants (22/22, 100%) reported that researchers can request data to be exported from the TRE at any time; however, the frequency of requests varied significantly per project, with some requiring daily exports and others exporting only at the end of a project.

Participants were keen to explore how the review process of data export can be improved and automated to decrease the review team's workload. However, there was concern that more frequent exports increased the risk of data leakage. For example, 2 consecutive releases featuring a subgroup of 26 and 27 patients, respectively, will each be acceptable in isolation, but comparing the 2 subgroups discloses additional information about the additional patient in the second release.

#### Limitations

As the manual export checking process uses significant staff time, some TREs apply limits on the number of exports or charge projects more for frequent use. A participant explained that the volume of export requests allowed is related to the cost model of the tenancy. For example, a TRE allowed only 2 releases for MSc or BSc projects.

#### Recommendations

Owing to the different types of data used across the different TREs and the different types of projects it is evident that there is no "one-size-fits-all" solution, but rather a solution needs to be sufficiently flexible to facilitate these differences between projects, data sets, and TREs. Automation can help to address these resource concerns and increase the speed and frequency at which researchers are able to export data. Although human checks are useful, the process has limitations and the risk of human error.

### *Potential Gaps in Export Checks*

#### Overview

Participants were asked whether they perceived any gaps in the export process and how they thought it can be improved. Of the 22 participants, 15 (68%) were not aware of any gaps or security concerns. The following concerns were raised by the other participants:

1. Researchers can be creative in finding a way to remove data, for example, by using screen capture to exfiltrate data, which will be difficult to detect.
2. Manual checks have the potential for human error.
3. Owing to the variety of data types that will be requested for export, it was difficult to find software that had the functionality to check all the file types. This variety also makes it challenging to bring together a review team with knowledge of where data may be accidentally or deliberately hidden, particularly for novel data types. None of the TREs were aware of any existing software tools that can be used for checking algorithmic data export requests.
4. Deficiencies in the audit trail make it impossible to see what the researchers have done in the TRE, because, sometimes, studies may have deviated from the original goal, and this was difficult to detect.

A participant mentioned that the manual process can be greatly enhanced by the following:

1. Effective training
2. Ensuring that staff rigorously check outputs
3. Applying the principles of appropriate frameworks, such as the Seven Safes, and nationally recognized "best practice" (eg, the Canadian essential requirements for operating data trusts [35])
4. Having a collaborative relationship with researchers throughout their project to mitigate and prevent malicious behavior

#### Recommendation

Typically, the 2 different models for output checking are principle-based and rule-based models [36,37]. The principle-based model better ensures confidentiality protection and utility of outputs from a statistical perspective, at the expense of being hard to standardize, automate, and verify.

## Safe Settings

### *Overview*

The safe setting controls cover the infrastructure and associated security measures that should be adopted by TREs. These controls specify that computing power and operating systems should enable a safe setting to sustain both economical scalability of compute for analysis (eg, images and genomics) and integral data security. Safe setting controls describe the best practices of policies, techniques, and security measures and strategies that are required when sharing data for analysis.

### *Computing Power and Operating System Offered to a "Standard" User*

#### Overview

Generally, TREs take the form of a virtual desktop infrastructure—each user receives remote access to a desktop environment with access to their project's data and appropriate software to analyze that data. Most TREs provide each user with their own virtual machine (VM), with fixed resources (particularly memory and processing) isolated from other projects and users, whereas a few TREs share a multiuser system more directly (known as "session-based" virtual desktop infrastructure), allowing a user to exploit the full hardware capacity of the host system when needed, at the expense of reduced isolation between users and projects.

Alternative approaches also exist; for example, the OpenSAFELY platform [38] provides indirect access to patient data. Rather than manipulating the data interactively, researchers develop analysis scripts against synthetic dummy data and then submit those scripts for remote execution. None of the groups that provide access via such alternative approaches agreed to participate in this study; therefore, they were not included in the analysis.

Most respondents (18/20, 82%) indicated that standard templates were usually used for the TRE and the computing power offered for a project will depend on the number of users who need to access it. Of the 22 participants, 1 (5%) creates custom configurations for every project, 2 (9%) have no flexibility available, and 12 (55%) reported that their TREs can scale-up depending on the researcher's requirements. Heavy compute usage will have high costs associated with it, which can be a barrier for many research projects. A participant mentioned that the maximum computing power configurations depend on each individual project's budget constraints. Table 1 lists the different computer power available across the TREs (some of the participants were not able to answer this question and some use the public cloud, so that resources are effectively limited only by budget).

A total of 53% (10/19) of the TREs reported that Windows (including Windows 10, Windows Server 2012, Windows Server 2019) was the standard build operating system. Of the 19 TREs, 4 (21%) responded that they can provide both Windows and Linux based on the researcher's request. In a TRE, Ubuntu was the only standard build. From the participants' responses, it was evident that there was great variety in the specifications available, with some having multiple orders of magnitude more capacity than others.

**Table 1.** Available computing power.

| Processing power and RAM | | Storage space | Allocation |
|---|---|---|---|
| 1 CPU[a] | 8 GB | 5 TB | VM[b] |
| 2 CPUs | 8 GB | 250 GB (fast scratch) | VM |
| 4 CPUs | 16 GB | —[c] | VM |
| 1.5 cores | 18 GB | 1 TB | VM |
| 4 to 8 cores | 32 to 64 GB | 60 to 80 GB | VM |
| 4 to 64 cores | — | 8 GB to 2 TB | Host |
| 16 cores | — | 96 GB | Host |
| **Dual Xeon processor** | | | |
| | — | 120 GB | VM |
| | ~2000 cores | — | Host |
| **GPU [d] cluster** | | | |
| | — | 200 TB | Host |
| | 4 TB | 32 TB | Host |

[a] CPU: central processing unit.

[b] VM: virtual machine.

[c] Information not available or no predetermined value.

[d] GPU: graphics processing unit.

### Limitations

Although most current researcher needs are met by existing TREs, it is clear from Table 1 that some TREs can find it challenging to support processor-intensive projects. Furthermore, most, but not all, of the TREs provide each project its own isolated VM, which can have implications for isolation and pose an increased security risk if malicious code was able to run and potentially access other projects and their data within a shared system as opposed to a project-specific VM.

### Recommendations

TREs should consider the scalability of their infrastructure to support resource-intensive projects in the future. Use of public cloud infrastructure enables much greater flexibility, for a price, and incorporates robust isolation between VMs as standard.

## *Data Security Measures Used in TREs to Mitigate the Risk of Unauthorized Access, Data Loss, and Misuse by Researcher*

### Unauthorized Access

The participants discussed different measures that were implemented to help prevent unauthorized access. Different controls were implemented across the participating TREs, depending on the underlying infrastructure. We present a full list of the controls that were discussed during the interview; however, not all the TREs implemented the full list of controls described in this paper:

1. Best practice password policy (which will include lockout after 2 or 3 incorrect attempts)
2. Access controls
3. Access to TRE only permitted via white-listed IP addresses
4. Fully automated account management
5. Sensitive projects may have restrictions on the location of the researcher (in its strictest form, this can include

permitting access only from a specific room [on campus] and via managed devices [restricted machines], or more generally, permitting IP addresses only from particular countries)

6. Manage file access
7. Active Directory hierarchical privileges
8. Session recording
9. Monitoring or audit system, such as IBM Guardium, SIEM, and Splunk
10. Multifactor authentication
11. Network segmentation
12. Compartmentalization to limit access to information to entities on a need-to-know basis to perform certain tasks, to protect information vendor firewalls (3 different vendors)
13. Patch management
14. Biannual pen testing

#### Data Loss

In the TREs, internet access is blocked, and users have limited access rights. The remote access is designed to prevent moving data in and out of the environment, except via the official channels, with appropriate controls in place: virtual hardware ports and copying data to the client system's clipboard are disabled. Some TREs also take steps to impede pasting; however, this is not reliably achievable (the direct paste shortcuts can be disabled, but it is trivial for a knowledgeable user to bypass this with a single command on client systems). Measures are also established to detect attempts to export data via other routes. Antimalware or antiransomware software and data loss prevention software are used.

#### Misuse by Researcher

The main countermeasure to misuse by researchers is training. Generally, this reinforces key principles to ensure that researchers understood their responsibilities and what activities are permitted or not permitted within the environment. Other significant mitigation strategies are checking the outputs and reviewing the project's scope. In 11% (2/19) of the TREs, researchers must be accredited by a particular organization before they are granted access to the environment (this accreditation requires the researchers to prove that they have appropriate qualifications and experience). Furthermore, researchers must sign an investigator's declaration stating that they will not misuse the environment, and the line managers and organizations will be held accountable if a user attempts to do anything malicious.

A TRE uses session recordings to help detect misuse. In this TRE, researchers' behavior such as keystrokes can be monitored. Another TRE uses a monitoring program from Darktrace to detect a user running a tool on their laptop to take screenshots [39]. Other 2 TREs have a full audit log from log-on to log-off, and another TRE plans to log activities to enable reconstruction in the event of a breach.

Many other controls were discussed, which included the following:

1. Researchers are not granted admin access in the TRE.
2. Researchers have access only to their own project's TRE storage.
3. Printing, mapping drives, and accessing external drives are not allowed.
4. Command prompt access is disabled.
5. ISO27001 policy rules via a cloud security posture management system.

#### Recommendations

The previously mentioned examples of current practices to detect and prevent instances of unauthorized access, data loss, and researcher misuse should be considered by all TREs to further improve security, where appropriate for the specific TRE infrastructure. Furthermore, TREs must have a legal agreement constraining access and use as their data security measure to mitigate the risk of misuse by the researcher. Programs to monitor and record researchers' behavior are also useful to reduce misuse.

### *Importing of Data or Code*

#### Overview

Participants were asked if they allow researchers to import data or code (including libraries) into the environment and, if so, what security measures (eg, software) are used to support this process:

1. A total of 63% (12/19) of the TREs allow the import of both code and data.
2. A total of 16% (3/19) of the TREs allow code (with some restrictions), but not external data sets.
3. A total of 11% (2/19) of the TREs allow data, but not code.
4. A total of 5% (1/19) of the TREs allow neither.

The import of data or code is subject to gatekeeper approval with a check that the import does not contain hidden data and that the code does not pose a threat to the security of the TRE. This gatekeeper approval process varies between TREs, but typically involves manual checks. In addition to scrutinizing the security risk posed by the data or code, this process can also involve checking the file size, file type, magic numbers, and known suffixes. In general, this process is supported by virus scans, static code analysis tools, and sample code execution in a sandboxed environment.

Some participants discussed the important role of "trust," and how training the researchers and trusting that they have no malicious intent is sufficient, based on the low risk of potential damage from malicious code and subject to low sensitivity of the data (refer to the Safe People section for more detail). Finally, a participant mentioned the role of monitoring to detect any malicious behavior, so that inappropriate or malfunctioning software can be identified.

#### Limitations

There was substantial reliance on manual checks to support this process. Furthermore, participants had clear concerns about the security implications of importing malicious code. The main concerns regarding the process of supporting code or data egress were highlighted as follows:

1. Ensuring that the AI algorithms or software imported into the environment do not include sensitive data.

2. It can be extremely time-consuming for the TRE staff or researchers to manually import code after each small change.

### Recommendations

Some of these security concerns can be mitigated by isolating each project within the TRE to minimize potential damage and limiting the privileges of the researchers in the environment, using virtualization or containerization techniques. Similar to the recommendations for data checks, there is a clear need for tools that can support the TRE team in checking the data and code that researchers wish to bring into the environment. Although there are clear concerns about fully automating this process, developing tools to support these checks can significantly speed up the process and assist with the detection of malicious code.

## *Support for Federated Queries of Data From External Sources*

### Overview

A total of 79% (15/19) of the TREs did not currently support federated queries from external sources, whereas the remaining TREs (4/19, 21%) confirmed that their TRE supported this. A participant described how their TRE can support federated queries via an integration tool on the Health and Social Care Network, using application programming interfaces, XDS.b cross-enterprise document sharing [40], and the Image Exchange Portal (Sectra AB) for imaging.

A federated query entails sending a request to ≥1 external data sources and receiving only the results of that query, as opposed to the usual pattern of the operators copying an entire data set into the TRE for later analysis. This can be more efficient—avoiding the need to prescreen all data on ingress to the TRE and allowing use of live data sources rather than static copies—and allowing easy aggregation of multiple sources; however, it is hard to implement securely and efficiently.

### Limitations

Federated queries are difficult to support while maintaining effective privacy and security controls, and they are not currently available in most TREs.

### Recommendations

Federated queries enable federated learning that can train machine learning (ML) algorithms from diverse data sets without exchanging data. Federated learning can be effective in diagnosing uncommon diseases, and it can also reinforce data privacy and security if the process of data being stored and processed is supported by privacy-preserving and cryptographic techniques [41-43]. Furthermore, federated learning complies with data protection regulations including General Data Protection Regulation (GDPR). However, federated learning is vulnerable to different attacks such as inference attacks (eg, membership and reconstruction attacks) [44] and poisoning attacks [45,46], which can violate GDPR. The possibility of some of these attacks can be mitigated by the application of privacy-preserving mechanisms including secure multiparty computation, differential privacy, and encrypted transfer learning methods [47]; however, differencing attacks remain difficult to guard against [48]. Supporting federated queries of data from external sources is a feature of interest for the next-generation TREs.

## *Audit and Workflow Management*

### Overview

Audit and monitoring are key aspects of a TRE. Many participants reported that they use project management tools to automate functions such as Jira [49], which can be customized to record transitions, such as a request being made by a user, review of the data to be exported, and subsequent acceptance or rejection of the export. Most TREs (13/19, 68%) reported that they keep a copy of the exported data.

The level of automation and functionality of auditing differs between TREs. The state of the art includes the following:

1. Real-time alerting on the digital airlock, providing a verbose description of user activity. The reports and alerts generated from this provides the IP address of the user and their username, along with the time, date, file name, file size, and few other supplementary fields.
2. All the activities in the TRE are logged, dashboards are used to support the monitoring of the activities, and reports are automatically generated.
3. If a user attempted to export data that was not permitted, this will be logged. If abnormal patterns are observed, the antimalware software (eg, Sophos plus quest tools [50]) will trigger alerts and log tickets on the system. The technical team and data owner will receive an email alert advising them that abnormal patterns had been detected.

### Limitations

Many TREs have little or no automation and automated auditing in place, thus limiting the available reporting and operational insights.

### Recommendations

Incorporating a logging and monitoring system into the TRE is important. This system can include log-in attempts, including username, time and date of access, IP address, type of activity conducted during the session (eg, which tools were used, for how long, and any processes that were running), details of any imports and exports (including file name and file size), and access type (successful or denied). Furthermore, having a real-time alert system can warn the TRE team promptly in case of any malicious attempts and assist in preventing unwanted disclosure and blocking access.

## Safe Computing

Participants were asked whether their TRE uses private (on-premises) or public cloud infrastructure ("Infrastructure as a service," such as Amazon's Amazon Web Services or Microsoft's Azure). In total, 64% (14/22) of the participants reported that their TREs use a private cloud. There were some concerns from these participants that data governance restrictions may make switching to a public cloud difficult. In all, 21% (4/19) of the TREs were already hosted in public clouds, and 10% (2/19) of the participants reported that they aim to switch to a public cloud in the future. Although costs are generally high in public clouds, the extra functionality and flexibility make this an attractive option when possible.

In comparison with local ("on-premise") servers, storing data, especially sensitive data (eg, health data), in the public cloud can be considered more secure because of its enhanced security measures and the expert advice available from the cloud service providers. However, most TREs currently use on-premise servers, which simplify data governance and data locality issues and provide consistent performance and predictable costs. It was interesting to see that some TREs plan to switch to public cloud in the future, as the public cloud can scale more easily and provide access to next-generation services such as AI or ML, containers, and so on.

## Safe People

The safe people controls are measures and policies to ensure that trusted researchers will use the platform in an appropriate manner.

### *Controls on the People Who Use the TRE*

#### Overview

Best practices for ensuring that the researchers accessing the environment are trustworthy and understand the importance of correct use of the TRE include the following: signing legal documents to agree that a researcher will avoid attempting to reidentify any individual, rapid disclosure of any vulnerabilities detected by a researcher, keeping log-in credentials private, and notification to the TRE if a researcher was leaving their institution. A participant reported that financial penalties can be a useful deterrent to misuse.

Moreover, 85% (17/20) of the participants responded that researchers using their TREs are required to complete training. This training typically consists of information governance, GDPR, awareness of issues related to privacy, ethics, security, information security, Medical Research Council training, and statistical disclosure control. Researchers are typically required to complete the training annually or before the beginning of each project. The nature of this training and subsequent contract or terms of use are typically determined by the data owner. For example, government security clearance is requested by the Defence Science and Technology Laboratory for access to their data.

In 79% (15/22) of TREs, researchers sign an agreement not to misuse the environment or the data. This agreement is also signed by a senior member within each organization. A participant stated that if a researcher is a student, a supervisor also needs to sign the agreement. There was a range of penalties applied across the TREs for violating the user agreement, which, in the most extreme form, can result in job loss, disciplinary measures, or, in some cases, compulsory retraining. Project approval is also required by the relevant data controller, and in some cases, the project also has to be signed by an ethics committee. According to a TRE, conditions specified in a nondisclosure agreement or access request form will impose constraints regarding appropriate use of the data and can pass all responsibilities for ensuring that the data were being used correctly on to the sponsoring organization.

#### Recommendations

Training, such as information governance training, is vital to ensure that researchers understand their responsibilities and should be considered by all TREs. Through training, researchers will clearly understand what they are allowed to do with the data. TREs must implement suitable review and management processes to further ensure that researchers are using the TREs appropriately. Best practice approaches to delivering effective training for TRE researchers, which not only support the use of the TRE but also facilitate building shared attitudes and responsibilities in protecting data, are widely available [36,37,51].

### *Controlling Access to the Environment for Trusted Users Only*

Overall, 76% (14/22) of the participants stated that access to their TREs is limited to those researchers who are associated with an approved (trusted) organization. Furthermore, 14% (2/14) of these participants stated that access was limited to organizations in the same country as the TRE, as specified by the data custodian. For 7% (1/14) of them, commercial organizations were not allowed to access the TRE under any circumstances. For other TREs that permit commercial organizations to access the environment, the criteria for approving these organizations were generally set higher than those for other organizations (eg, universities). In a TRE, although requests from commercial organizations were considered, they needed a university sponsor or health sponsor to be approved. Another participant responded that commercial customers did not need to be associated with an academic institution. In this case, a review committee determines which projects will be approved for commercial customers.

In a TRE, access is granted only to the users of their own university. In this TRE, an external visitor account will be granted access only if the visitor was sponsored by university staff. In another TRE, researchers can access the environment from a university or NHS-based organization (ie, using white-listed IP addresses). A TRE adopted additional restrictions for the researchers, for example, ensuring that access was permitted only from a safe room or that the device used to access the TRE was a managed device and not a personal device. In this case, these restrictions were set by the data controller.

## Participants' Recommendations for TRE Enhancements

In all, 27% (6/22) of the participants indicated that they would like improved support for programming capabilities in the environment (eg, Python and R), to advance the analytical capabilities of their TRE and subsequently support large-scale studies. Support for importing data, algorithms, and code to the environment was frequently described as another high-priority feature. However, the licensing of proprietary software tools presents a further limitation regarding the incorporation of software into the environment, because not all licenses cover use within a TRE.

Overall, 9% (2/22) of the participants confirmed that they would like to support federated learning to advance data movement among TREs, where data sets need to be shared and accessible.

Support for additional data modalities, such as imaging and genomic data, needs to follow a proper risk assessment, and TREs would have to ensure that they liaise with data custodians regarding the specific risks. It was widely acknowledged by the participants that there were many security challenges around allowing researchers to bring their own data and code into the environment, and until solutions to these challenges have been developed, many TREs will be reluctant to support this.

A total of 23% (5/22) of the participants indicated that they would like to simplify the process for researchers to access data within their TRE in the future. The process and checks required before researchers are granted access to the data were perceived as cumbersome and slow. Sometimes, this administrative process is further delayed owing to backlog of project review requests, committees being slow to make decisions, ethics board approval, and researchers completing relevant training courses and privacy training. Researchers are eager to have access to the TRE and its data promptly; hence, TREs desired to simplify this process. Overall, 36% (8/22) of the participants discussed how they would like to improve the governance processes. A participant stated that all data sets in their TRE were treated as high risk and had to go through the same governance process, even though some data sets were actually low risk. A participant suggested that it will be useful to conduct a national risk-benefit analysis of sharing standardized data sets for research. The participant acknowledged that there was no systematic approach to review data sets to determine if there were certain conditions under which these can be used by researchers without the full governance checks.

Some TREs are considering migrating to a public cloud for improved scalability and flexibility, including GPU access, great on-demand computing power, and reduced management overheads, whereas several TREs have already made this transition.

A TRE is looking to enhance their security through improved logging of activities, such as data copying between machines, and better behavior tracking.

Finally, a participant discussed concerns about intellectual property when the code was developed within the TRE. The participant acknowledged that researchers may have concerns regarding how the code that they develop or test in a TRE can be accessed by the TRE operators. Policies and practices to govern this should be established to protect both parties. Technical solutions to this problem, such as trusted computing and enclave approaches, can also be explored.

## Discussion

This study reviewed the existing controls used by UK and international TREs that participated in our structured interviews. These controls cover a subset of the seven "Safes," comprising safe people, safe setting, safe computing (an extension of safe setting), safe data, and safe outputs. The features that most need further work for next-generation TREs are the following:

1. Advancing analytical power (high-performance computing clusters and GPUs) available within the environment to support large-scale studies
2. Bringing data, algorithms, or code into the environment and addressing the security challenges arising from this
3. Being able to develop ML and AI algorithms within the TRE and export them
4. Supporting federated queries of data from external sources
5. Supporting additional data modalities such as imaging and genomic data
6. Simplifying the process of accessing data for researchers
7. Scalability

This study analyzed the extent to which TREs can support the import and export of different data types. The process used is largely manual, with some TREs using software to support this process. Finding suitable software to support the automation of the data center process was identified as a key priority for most TREs. Furthermore, the application of ML techniques in TREs can be useful for predicting the malicious use of accessed data by researchers. It was evident that most TREs do not have specific tools to manage and mitigate the potential risk of reidentification, and they rely on analysts' knowledge and judgment and communication with the data controller.

There is a lack of support for AI and ML development in TREs, and there is a concern that researchers can perform malicious activities owing to the AI and ML structure, for example, exporting sensitive data that can be vulnerable to exposure following attacks against the AI model or overtraining the AI algorithm. The difficulties in detecting these exports were acknowledged as a significant challenge by the participants.

The computing power available to researchers is generally adequate for current needs (observational studies using statistical analysis tools); however, there was a clear desire to ensure that this was scalable to meet researcher's requirements for analyzing big data and for AI development. Some TREs already appear to be significantly constrained. There is significant variety in the extent of the security measures used to mitigate the risk of unauthorized access, data loss, and misuse by the researcher, and there are some concerns regarding the implications of next-generation capabilities on the security of the TRE and protecting the data. Furthermore, there is a need for advanced information governance for TREs encompassing incoming and outgoing automated data feeds, ad hoc incoming data and algorithms, and ad hoc outgoing data and algorithms. Finding appropriate solutions to meet these needs should be explored in future studies.

Alternative approaches to the remote desktop approach exist, and further exploration of the relative merits of these alternatives will be valuable, particularly in the context of new and evolving types of data.

## Acknowledgments

This project was supported by the UK Medical Research Council and Engineering & Physical Sciences Research Council PICTURES programme (grant number MR/S010351/1).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview questions.
[DOCX File , 51 KB - jmir_v24i9e33720_app1.docx ]

## References

1. A charter for safe havens in Scotland. The Scottish Government. URL: https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/ [accessed 2021-09-15]
2. Hubbard T, Reilly G, Varma S, Seymour D. Trusted Research Environments (TRE) Green Paper. Zenodo. 2020 Jul 21. URL: https://zenodo.org/record/4594704#.Ys7RuHZBzIU [accessed 2022-09-06]
3. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC, US: The National Academies Press; 2015.
4. Lea NC, Nicholls J, Dobbs C, Sethi N, Cunningham J, Ainsworth J, et al. Data safe havens and trust: toward a common understanding of trusted research platforms for governing secure and ethical health research. JMIR Med Inform 2016 Jun 21;4(2):e22 [FREE Full text] [doi: 10.2196/medinform.5571] [Medline: 27329087]
5. HIC Trusted Research Environment. University of Dundee. URL: https://www.dundee.ac.uk/hic/hicsafehaven/ [accessed 2021-09-16]
6. eDRIS: enabling research access to Scottish health datasets. Edinburgh International Data Facility. URL: https://www.ed.ac.uk/edinburgh-international-data-facility/updates-events/electronic-data-research-and-innovation-service [accessed 2021-09-16]
7. Jones K, Ford D, Thompson S, Lyons R. A profile of the SAIL databank on the UK secure research platform. Int J Popul Data Sci 2019 Nov 20;4(2):1134 [FREE Full text] [doi: 10.23889/ijpds.v4i2.1134] [Medline: 34095541]
8. Jones K, Ford D, Ellwood-Thompson S, Lyons RA. The UK Secure eResearch Platform for public health research: a case study. Lancet 2016 Nov;388:S62 [FREE Full text] [doi: 10.1016/s0140-6736(16)32298-x]
9. Office for national statistics. ADR UK. URL: https://www.adruk.org/about-us/our-partnership/office-for-national-statistics/ [accessed 2021-09-16]
10. Administrative data is invaluable resource for public good let's use it. ADR UK. URL: https://www.adruk.org/ [accessed 2021-09-16]
11. Access levels and conditions. UK Data Service. URL: https://www.ukdataservice.ac.uk/use-data/secure-lab.aspx [accessed 2021-09-16]
12. Desai T, Ritchie F, Welpton R. Five Safes: designing data access for research. University of the West of England. URL: https://www2.uwe.ac.uk/faculties/bbs/Documents/1601.pdf [accessed 2020-04-06]
13. El Emam K, Dankar FK. Protecting privacy using k-anonymity. J Am Med Inform Assoc 2008;15(5):627-637. [doi: 10.1197/jamia.M2716] [Medline: 18579830]
14. Templ M, Kowarik A, Meindl B. Statistical disclosure control for micro-data using the package. J Stat Soft 2015;67(4):1-36. [doi: 10.18637/jss.v067.i04]
15. Privitar homepage. PRIVITAR. URL: https://www.privitar.com/ [accessed 2021-06-10]
16. TriNetX, InSite unite to establish world’s largest clinical research network. TriNetX. 2019 Apr 2. URL: https://trinetx.com/insite/ [accessed 2021-09-16]
17. Privacy Analytics' PARAT 6.0, the next generation de-identification software, unlocks the potential for faster, better health data analytics. Health IT Outcomes. 2014 Nov 20. URL: https://www.healthitoutcomes.com/doc/privacy-analytics-parat-de-identification-software-0001 [accessed 2022-09-05]
18. De-identification. Future of Privacy Forum. URL: https://fpf.org/issue/deid/ [accessed 2021-09-16]
19. Hogue P. The risk of triangulation: you may just be a piece of the puzzle. Securityweek. URL: https://www.securityweek.com/risk-triangulation-you-may-just-be-piece-puzzle [accessed 2021-09-16]
20. Data anonymisation - a key enabler for clinical data sharing. European Medicines Agency. 2018 Dec 04. URL: https://www.ema.europa.eu/en/documents/report/report-data-anonymisation-key-enabler-clinical-data-sharing_en.pdf [accessed 2022-09-06]
21. Rigaki M, Garcia S. A survey of privacy attacks in machine learning. ArXiv 2021 [FREE Full text] [doi: 10.5860/choice.45-0602]

22. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. 2015 Presented at: CCS'15: The 22nd ACM Conference on Computer and Communications Security; Oct 12 - 16, 2015; Denver Colorado USA URL: https://doi.org/10.1145/2810103.2813677 [doi: 10.1145/2810103.2813677]
23. Nind T, Sutherland J, McAllister G, Hardy D, Hume A, MacLeod R, et al. An extensible big data software architecture managing a research resource of real-world clinical radiology data linked to other health data from the whole Scottish population. Gigascience 2020 Sep 29;9(10):giaa095 [FREE Full text] [doi: 10.1093/gigascience/giaa095] [Medline: 32990744]
24. Nicolae MI, Sinn M, Tran MN, Buesser B, Rawat A, Wistuba M, et al. Adversarial robustness toolbox v1.0.0. arXiv 2019 [FREE Full text]
25. Liu Y, Wen R, He X, Salem A, Zhang Z, Backes M, et al. ML-Doctor: holistic risk assessment of inference attacks against machine learning models. arXiv 2021 [FREE Full text] [doi: 10.14722/ndss.2019.23119]
26. Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. In: Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP). 2017 Presented at: 2017 IEEE Symposium on Security and Privacy (SP); May 22-26, 2017; San Jose, CA, USA URL: https://ieeexplore.ieee.org/document/7958568 [doi: 10.1109/sp.2017.41]
27. Salem A, Zhang Y, Humbert M, Berrang P, Fritz M, Backes M. ML-leaks: model and data independent membership inference attacks and defenses on machine learning models. In: Proceedings of the Network and Distributed Systems Security (NDSS) Symposium 2019. 2019 Presented at: Network and Distributed Systems Security (NDSS) Symposium 2019; Feb 24-27, 2019; San Diego, CA, USA URL: https://www.ndss-symposium.org/wp-content/uploads/2019/02/ndss2019_03A-1_Salem_paper.pdf [doi: 10.14722/ndss.2019.23119]
28. Qian J, Li X, Zhang C, Chen L. De-anonymizing social networks and inferring private attributes using knowledge graphs. In: Proceedings of the IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications. 2016 Presented at: IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications; Apr 10-14, 2016; San Francisco, CA, USA URL: https://ieeexplore.ieee.org/document/7524578 [doi: 10.1109/infocom.2016.7524578]
29. Al-Rubaie M, Chang JM. Reconstruction attacks against mobile-based continuous authentication systems in the cloud. IEEE Trans Inform Forensic Secur 2016 Dec;11(12):2648-2663 [FREE Full text] [doi: 10.1109/tifs.2016.2594132]
30. Takemura T, Yanai N, Fujiwara T. Model extraction attacks on recurrent neural networks. J Inform Process 2020;28:1010-1024 [FREE Full text] [doi: 10.2197/ipsjjip.28.1010]
31. Reith R, Schneider T, Tkachenko O. Efficiently stealing your machine learning models. In: Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society. 2019 Presented at: CCS '19: 2019 ACM SIGSAC Conference on Computer and Communications Security; Nov 11, 2019; London United Kingdom URL: https://dl.acm.org/doi/10.1145/3338498.3358646 [doi: 10.1145/3338498.3358646]
32. Zhang Y, Jia R, Pei H, Wang W, Li B, Song D. The secret revealer: generative model-inversion attacks against deep neural networks. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020 Presented at: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); Jun 13-19, 2020; Seattle, WA, USA URL: https://ieeexplore.ieee.org/document/9156705 [doi: 10.1109/cvpr42600.2020.00033]
33. Aslanyan Z, Vasilikos P. Privacy-preserving machine learning. Alexandra Institute. 2020 Oct. URL: https://tinyurl.com/yck5swm7 [accessed 2022-09-06]
34. Macrae C. Governing the safety of artificial intelligence in healthcare. BMJ Qual Saf 2019 Jun 12;28(6):495-498 [FREE Full text] [doi: 10.1136/bmjqs-2019-009484] [Medline: 30979783]
35. Paprica P, Sutherland E, Smith A, Brudno M, Cartagena R, Crichlow M, et al. Essential requirements for establishing and operating data trusts: practical guidance co-developed by representatives from fifteen Canadian organizations and initiatives. Int J Popul Data Sci 2020 Aug 24;5(1):1353 [FREE Full text] [doi: 10.23889/ijpds.v5i1.1353] [Medline: 33644412]
36. Alves K, Ritchie F. Runners, repeaters, strangers and aliens: operationalising efficient output disclosure control. Statistic J IAOS 2020 Nov 25;36(4):1281-1293. [doi: 10.3233/sji-200661]
37. Brandt M, Franconi L, Guerke C, Hundepool A, Lucarelli M, Mol J, et al. Guidelines for the checking of output based on microdata research. University of the West of England. URL: https://uwe-repository.worktribe.com/output/983615/guidelines-for-the-checking-of-output-based-on-microdata-research [accessed 2022-09-08]
38. Williamson E, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug;584(7821):430-436 [FREE Full text] [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
39. Featured resources. Darktrace. URL: https://www.darktrace.com/en/resources/ [accessed 2021-09-16]
40. Cross-Enterprise Document Sharing (XDS.b). Integrating the Healthcare Enterprise. URL: https://profiles.ihe.net/ITI/TF/Volume1/ch-10.html [accessed 2022-09-06]
41. Phong LT, Aono Y, Hayashi T, Wang L, Moriai S. Privacy-preserving deep learning via additively homomorphic encryption. IEEE Trans Inform Forensic Secur 2018 May;13(5):1333-1345. [doi: 10.1109/tifs.2017.2787987]

42. Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, et al. Federated learning with differential privacy: algorithms and performance analysis. IEEE Trans Inform Forensic Secur 2020;15:3454-3469. [doi: 10.1109/tifs.2020.2988575]
43. Truong N, Sun K, Wang S, Guitton F, Guo Y. Privacy preservation in federated learning: an insightful survey from the GDPR perspective. Comput Security 2021 Nov;110:102402. [doi: 10.1016/j.cose.2021.102402]
44. Dwork C, Smith A, Steinke T, Ullman J. Exposed! A survey of attacks on private data. Annu Rev Stat Appl 2017 Mar 07;4(1):61-84. [doi: 10.1146/annurev-statistics-060116-054123]
45. Bhagoji AN, Chakraborty S, Mittal P, Calo S. Analyzing federated learning through an adversarial lens. In: Proceedings of the 36th International Conference on Machine Learning. 2019 Presented at: 36th International Conference on Machine Learning; Jun 09-15, 2019; Long Beach, California, USA URL: https://arxiv.org/abs/1811.12470 [doi: 10.48550/arXiv.1811.12470]
46. Bagdasaryan E, Veit A, Hua Y, Estrin D, Shmatikov V. How to backdoor federated learning. In: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistic. 2020 Presented at: Twenty Third International Conference on Artificial Intelligence and Statistic; Aug 26-28, 2020; Online URL: https://arxiv.org/abs/1807.00459 [doi: 10.48550/arXiv.1807.00459]
47. Salem M, Taheri S, Yuan J. Utilizing transfer learning and homomorphic encryption in a privacy preserving and secure biometric recognition system. Computers 2018 Dec 29;8(1):3 [FREE Full text] [doi: 10.3390/computers8010003]
48. Green E, Ritchie F, Tava F, Ashford W, Ferrer Breda P. The present and future of confidential microdata access: post-workshop report. University of the West of England. 2021 Nov 25. URL: https://uwe-repository.worktribe.com/output/8175728 [accessed 2022-09-06]
49. Jira Software. ATLASSIAN. URL: https://www.atlassian.com/software/jira [accessed 2021-09-16]
50. Sophos Cybersecurity as a Service. URL: https://www.sophos.com/ [accessed 2022-09-06]
51. Ritchie F, Green E. Data access project: final report. Australian Department of Social Services. 2016 Sep 06. URL: http://eprints.uwe.ac.uk/31874/ [accessed 2022-09-06]

## Abbreviations

**AI:** artificial intelligence
**GDPR:** General Data Protection Regulation
**GPU:** graphics processing unit
**ML:** machine learning
**NHS:** National Health Service
**TRE:** Trusted Research Environment
**VM:** virtual machine

---

*Edited by R Kukafka; submitted 20.09.21; peer-reviewed by F Ritchie, M Mars; comments to author 29.10.21; revised version received 22.03.22; accepted 30.05.22; published 20.09.22.*

*Please cite as:*
*Kavianpour S, Sutherland J, Mansouri-Benssassi E, Coull N, Jefferson E*
*Next-Generation Capabilities in Trusted Research Environments: Interview Study*
*J Med Internet Res 2022;24(9):e33720*
*URL: https://www.jmir.org/2022/9/e33720*
*doi:10.2196/33720*
*PMID:36125859*

Original Paper

# Medical Staff and Resident Preferences for Using Deep Learning in Eye Disease Screening: Discrete Choice Experiment

Senlin Lin[1,2*], MSc; Liping Li[3*], MPH; Haidong Zou[1,2], MD; Yi Xu[1,2], MD; Lina Lu[1,2], MPH

[1]Shanghai Eye Disease Prevention and Treatment Center, Shanghai Eye Hospital, Shanghai, China

[2]Shanghai Key Laboratory of Ocular Fundus Diseases, Shanghai General Hospital, Shanghai Engineering Center for Visual Science and Photomedicine, Shanghai, China

[3]Shanghai Hongkou Center for Disease Control and Prevention, Shanghai, China

[*]these authors contributed equally

**Corresponding Author:**
Lina Lu, MPH
Shanghai Eye Disease Prevention and Treatment Center, Shanghai Eye Hospital
No. 1440, Hongqiao Road
Shanghai, 200336
China
Phone: 86 02162539696
Email: lulina781019@qq.com

## Abstract

**Background:** Deep learning–assisted eye disease diagnosis technology is increasingly applied in eye disease screening. However, no research has suggested the prerequisites for health care service providers and residents willing to use it.

**Objective:** The aim of this paper is to reveal the preferences of health care service providers and residents for using artificial intelligence (AI) in community-based eye disease screening, particularly their preference for accuracy.

**Methods:** Discrete choice experiments for health care providers and residents were conducted in Shanghai, China. In total, 34 medical institutions with adequate AI-assisted screening experience participated. A total of 39 medical staff and 318 residents were asked to answer the questionnaire and make a trade-off among alternative screening strategies with different attributes, including missed diagnosis rate, overdiagnosis rate, screening result feedback efficiency, level of ophthalmologist involvement, organizational form, cost, and screening result feedback form. Conditional logit models with the stepwise selection method were used to estimate the preferences.

**Results:** Medical staff preferred high accuracy: The specificity of deep learning models should be more than 90% (odds ratio [OR]=0.61 for 10% overdiagnosis; *P*<.001), which was much higher than the Food and Drug Administration standards. However, accuracy was not the residents' preference. Rather, they preferred to have the doctors involved in the screening process. In addition, when compared with a fully manual diagnosis, AI technology was more favored by the medical staff (OR=2.08 for semiautomated AI model and OR=2.39 for fully automated AI model; *P*<.001), while the residents were in disfavor of the AI technology without doctors' supervision (OR=0.24; *P*<.001).

**Conclusions:** Deep learning model under doctors' supervision is strongly recommended, and the specificity of the model should be more than 90%. In addition, digital transformation should help medical staff move away from heavy and repetitive work and spend more time on communicating with residents.

***(J Med Internet Res 2022;24(9):e40249)*** doi:10.2196/40249

**KEYWORDS**

discrete choice experiment; preference; artificial intelligence; AI; vision health; screening

## Introduction

Vision loss, defined as either visual impairment or blindness, is becoming a vital aspect of public health [1], affecting economic, educational, and employment opportunities, reducing the quality of life, and increasing the risk of death [1]. Therefore, according to the recent eye care competency framework by the World Health Organization, the continuum of eye care across all levels of the health system should be highlighted, particularly primary health care, to support universal health coverage [2].

High-quality eye disease prevention health care, such as effective screening, can help eliminate almost 57% of all blindness cases [3]. Nowadays, artificial intelligence (AI) is gradually adopted in eye disease screening and may assist in addressing the limited and difficult-to-sustain resources in screening capacity, personnel costs, and diagnosis expertise [4]. The accuracy of AI models greatly affects the cost-effectiveness of eye disease screening [5]. Unfortunately, though the US Food and Drug Administration (FDA) had set a mandatory level of accuracy with a sensitivity of more than 85% and a specificity of more than 82.5% [6], the accuracy of AI-assisted eye disease screening systems in the real world were far worse than that reported in the model development phase [7]. Therefore, it is essential to make clear the medical staff and resident requirements of the accuracy of AI models in the community-based eye disease screening in the real world. However, no related research has been conducted thus far.

To fill this evidence gap, we conducted discrete choice experiments (DCEs) for health care providers and residents in Shanghai, China, from August 2021 to January 2022. We aimed to reveal the preferences of medical staff and residents for using AI technology in community-based eye disease screening, particularly their preference for accuracy. The DCE technique, originating in mathematical psychology, has been introduced in health economics to elicit preferences for health and health care [8]. Additionally, the DCE technique is predictive of choices, mimicking real-world decisions in health care decision-making (correctly predicting >93% of choices) [9].

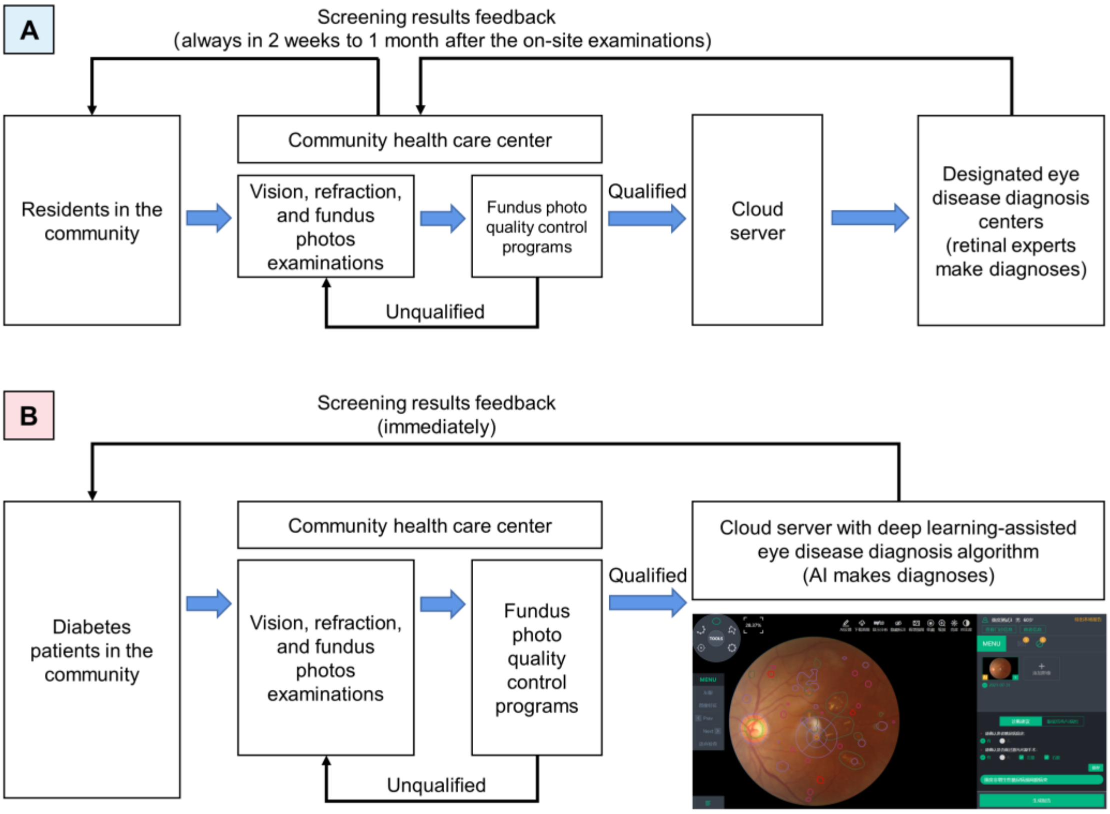
# Methods

## Study Setting

Shanghai, with a population of 24 million in 2019, is the economic, science, and technology innovation center in China. It is also one of the first cities in the world to adopt deep learning (DL) models to establish affordable and sustainable community-based eye disease screening systems. Since 2015, a teleophthalmology-based eye disease screening system covering all community health service centers has been developed in Shanghai. Residents can take free fundus photographs once a year by the trained general practitioners (GPs) in community health service centers. The fundus photos are then sent to the designated eye disease diagnosis centers through a dedicated information system. After the ophthalmologists in the diagnosis centers read the fundus photos and make diagnoses, the screening results are returned to the community health service centers. The GPs may inform residents of the screening results and provide medical advice.

In 2020, an AI-assisted eye disease screening system was established using DL model on cloud servers instead of ophthalmologists in the diagnosis centers making the screening diagnoses (Figure 1) [10-12]. The accuracy of DL models used for community-based eye disease screening has been reported widely [6,11,13,14]. Thus far, 56 community health service centers have shifted to the AI-assisted eye disease screening system. In 2021, these community health service centers screened over 40,000 residents with the help of the DL model and found over 7000 residents with suspected eye diseases.

**Figure 1.** Process of community-based eye disease screening in Shanghai. A: teleophthalmology-based eye disease screening system; B: deep learning–assisted eye disease screening system. The photograph in the lower right corner of process B is a sample of the operation interface of the deep learning–assisted eye disease diagnosis system. AI: artificial intelligence.



## Discrete Choice Experiments and Participants Inclusion

We conducted 2 DCEs to assess medical staff's preferences (experiment 1) and residents' preferences (experiment 2) for using the DL model in community-based eye disease screening. The main reason for using a DCE is that simply asking the respondents to rate the screening strategy attributes or choose their preferred item from a scale generally yields no more information than the fact that they want all the benefits and none of the indirect or direct costs [15]. Choosing between alternatives forces them to make a trade-off and choose, as in real life, between options that may increase utility (eg, improved diagnosis accuracy) and decrease utility (eg, screening cost of 40 CNY [US $6.15] per resident instead of being free).

Based on previously published literature [4-7,16-18], 4 attributes were identified initially to describe the outline of the community-based eye disease screening, including the accuracy, screening result feedback efficiency, level of ophthalmologist's involvement, and cost. It was worth stating that "screening result feedback efficiency" was included in the attributes because nearly instantaneous feedback might increase compliance [18]; moreover, "level of ophthalmologists involvement" was included because algorithmic aversion might exist [19]. To assess the appropriateness of these potential attributes and their levels, 5 experts on eye care were interviewed face-to-face in the Shanghai Eye Disease Control and Treatment Center. Based on these interviews, the attribute *accuracy* was divided into the following 2 attributes: "missed diagnosis rate" and "overdiagnosis rate," as they might have different impacts on the acceptability of eye disease screening. In addition, 2 new attributes were added: "organizational form" and "screening result feedback form," as the adoption of the DL model had the potential to reform the screening programs. As a result, 7 attributes were used to describe the outline of the community-based eye disease screening, and each attribute was divided into 3-6 levels (Table 1). Three SAS (SAS Institute Inc) procedures—"%mktruns," "%mktex," and "%choiceff"—were used to develop the questionnaire [20]. The questionnaire consisted of the following two parts: the respondent's basic information, such as sex and age, and a few choice sets, each of which contained 2 options with different screening attribute levels (Figure 2). The respondents were asked to choose the more favorable option in each choice set, and they were not allowed to choose both or neither in a set [21].

In Experiment 1, one municipal and 16 district-level eye disease control centers and over 250 community health service centers in Shanghai were enrolled. To receive rational rather than imaginary choices, the following two strict inclusion criteria were set: (1) they had over 5 years of experience in teleophthalmology-based eye disease screening and (2) they had over 1 year of experience in DL-assisted eye disease screening. A total of 34 institutions met the criteria, including 1 (3%) municipal, 16 (47%) district-level eye disease control centers, and 17 (50%) community health service centers (Figure 3). All the 40 key persons in charge of community-based eye disease screening in these 34 institutions were invited and agreed to participate in the experiment. Due to the limited number of respondents, we had to ask each one to answer a relatively large number of questions. According to the rule of thumb, as proposed by Johnson and Orme [22], we divided the alternative screening strategies into 30 choice sets of 2 options to ensure that the sample size of 40 people met the statistical requirements. The experiment was conducted in the form of a self-administered questionnaire, with a trained investigator on standby to interpret the questionnaire. One respondent quit because of temporary work arrangements. Therefore, data from 39 medical staff were available in the final analysis.

In Experiment 2, we randomly selected 2 from the 17 community health service centers involved in Experiment 1 and conducted the residents' investigation when carrying out the AI-assisted community-based eye disease screening. All the residents who participated in the screening were invited to the experiment. Because the number of residents was relatively large, we divided the alternative screening strategies into 10 choice sets of 2 options to reduce the response burden for each respondent. According to the rule of thumb, as proposed by Johnson and Orme [22], the minimum of the required sample size was 125. A total of 318 residents were investigated (Figure 3). To help the residents understand the questionnaire, the experiment was conducted using face-to-face questioning by trained investigator.

**Table 1.** Attributes and levels in the discrete choice experiments.

| Attributes | Levels | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **Performance expectancy** | | | | | | |
| Missed diagnosis rate (%) | None | 5 | 10 | 15 | 20 | —[a] |
| Overdiagnosis rate (%) | None | 5 | 10 | 15 | 20 | — |
| Screening result feedback efficiency | Immediately | In 2 weeks | In 1 month | — | — | — |
| **Effort expectancy** | | | | | | |
| Level of ophthalmologist involvement | Fully automated[b] DL[c] model | Semiautomated[d] DL model | Fully manual diagnosis[e] | — | — | — |
| **Facilitating conditions** | | | | | | |
| Organizational form | Centralized screening[f] | Residents' health self-examination cabin[g] | Opportunity screening in outpatient[h] | — | — | — |
| Cost | Free | 40 CNY[i] | 80 CNY | 120 CNY | 160 CNY | 200 CNY |
| Screening result feedback form | Screening results[j] | Screening results and medical advice[k] | Screening results, medical advice, and oral explanation by GP[l,m] | — | — | — |

[a] Not available.

[b] The screening results were provided entirely by the deep learning model, and the ophthalmologists were not involved in the diagnostic process.

[c] DL: deep learning.

[d] The deep learning model performed the initial screening of fundus photographs and then the ophthalmologists reviewed the results.

[e] The screening results were provided entirely by the ophthalmologists and the deep learning model was not involved in the diagnostic process.

[f] The community health service center informed the residents to undergo the screening at a uniform place and time.

[g] The equipment needed for screening was placed in a specific cabin in the community health service center, and residents could go to the cabin for self-examination at any time.

[h] Residents with chronic diseases and other risk factors would be recommended by general practitioners for eye disease screening during their outpatient follow-up.

[i] US 1$=6.5 CNY.

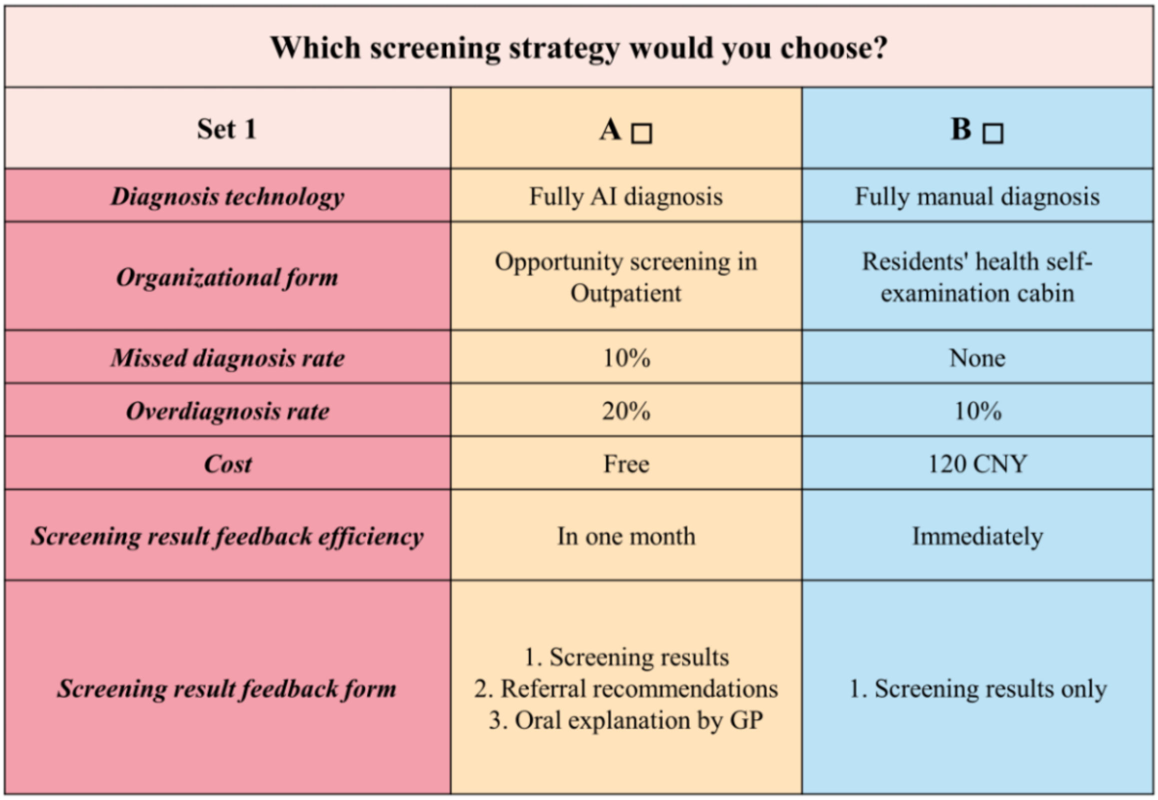
[j] The report with only the screening results would be given to the residents without any recommendations or explanations.

[k] The report with the screening results and referral recommendations would be given to the residents without explanations.

[l] Besides the report with the screening results and referral recommendations that would be given to the residents, a general practitioner would also explain the meaning of the report.

[m] GP: general practitioner.

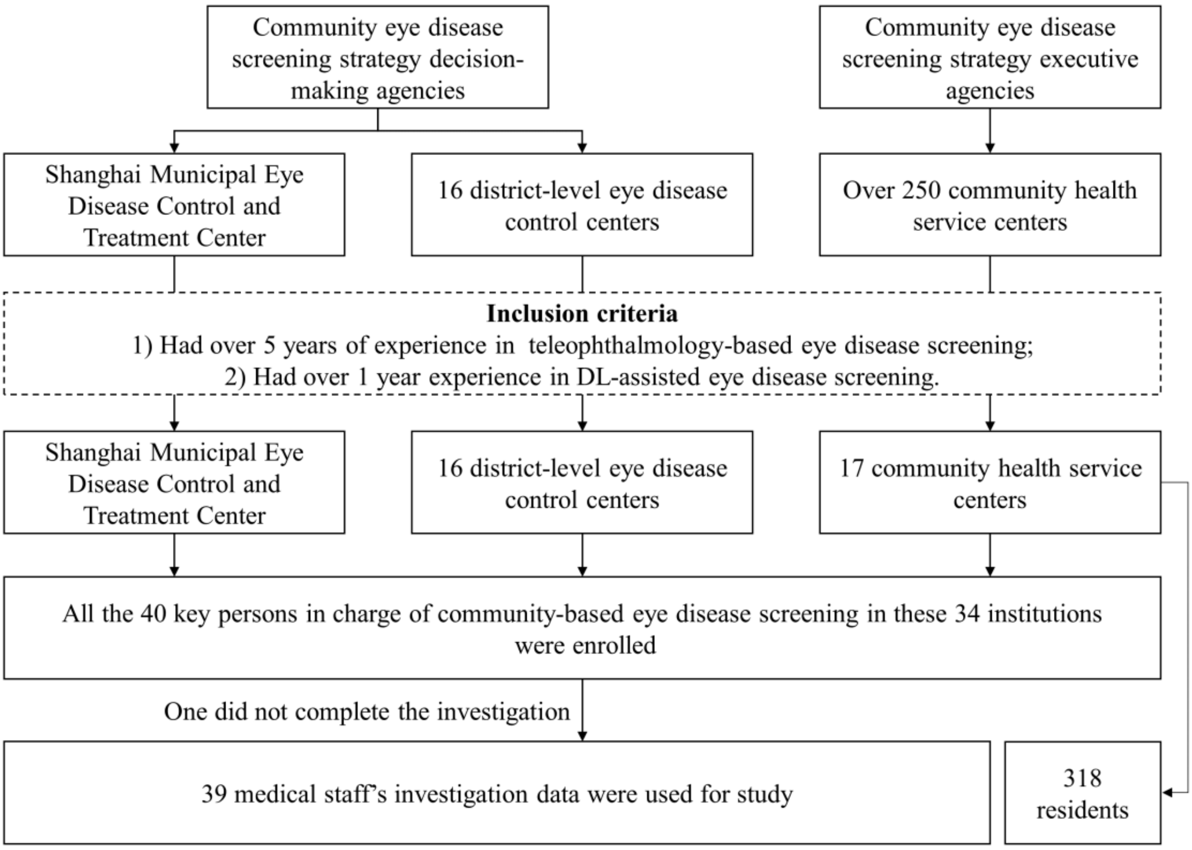
**Figure 2.** Example of the choice sets applied. Both options include the same 7 attributes. Health care service providers and residents were asked to decide between options A and B (in 2021, 1 USD=6.5 CNY). AI: artificial intelligence; GP: general practitioner.



**Which screening strategy would you choose?**

| Set 1 | A ☐ | B ☐ |
|---|---|---|
| ***Diagnosis technology*** | Fully AI diagnosis | Fully manual diagnosis |
| ***Organizational form*** | Opportunity screening in Outpatient | Residents' health self-examination cabin |
| ***Missed diagnosis rate*** | 10% | None |
| ***Overdiagnosis rate*** | 20% | 10% |
| ***Cost*** | Free | 120 CNY |
| ***Screening result feedback efficiency*** | In one month | Immediately |
| ***Screening result feedback form*** | 1. Screening results<br>2. Referral recommendations<br>3. Oral explanation by GP | 1. Screening results only |

**Figure 3.** Medical staff and residents inclusion process. DL: deep learning.



### Statistical Analyses

Mean, median, and standard deviation were calculated for the quantitative variables. For categorical variables, the number in a specific category was calculated as a percentage. Pearson chi-square test for nominal variables and Mann-Whitney *U* test for continuous variables were used for statistical analysis. Conditional logit models with the stepwise selection method were used to explore the significant preferences for each attribute level, with the choice responses as the binary dependent variable and the difference in levels for each of the attributes as the independent variables [21]. Two models were used to estimate the medical staff's and residents' preference respectively, expressed as odds ratios (ORs) for each attribute level. SAS 9.4 (SAS Institute Inc) were used for statistical analysis. The level of significance was set at *P*<.05.

### Ethics Approval

All participants were adults. Written informed consent from all participants was obtained before enrollment. The study adhered to the principles of the Declaration of Helsinki on Ethics. This study was approved by the Shanghai General Hospital Ethics Committee (2022SQ272).

## Results

The medical staff's mean age was 39.67 (SD 6.98) years, and they had been responsible for eye disease screening for 6.73 (SD 5.76) years on average. The residents' mean age was 68.62 (SD 6.96) years; Of the 318 participants, 120 (37.74%) were male and 198 (62.26%) were female. Detailed characteristics of the respondents are shown Table 2.

Table 3 presents the results of the conditional logit models, evaluating the influence of the tested attribute levels on medical staff's and residents' preferences. Among the 39 medical staff, the impact of selected attributes on preferences was statistically significant for 4 of the 7 attributes. Generally, medical staff prefer attribute levels with AI technology, lower overdiagnosis rates, lower screening costs, and higher screening result feedback efficiency. The results for the attribute "organizational form," "missed diagnosis rate," and "screening result feedback form" were inconclusive—none of the attribute levels were associated with statistically significant utility differences.

Further, we focused on the accuracy of the diagnosis. For the missed diagnosis rate, there were no significant differences of medical staff's preferences for a missed diagnosis rate between 0% and 20%. However, for the overdiagnosis rate, compared with no overdiagnosis, medical staff's preference for the 10% overdiagnosis rate significantly decreased (OR=0.61; *P*<.001).

Among the 318 residents, the influence of selected attributes on preferences was statistically significant for 3 of the 7 attributes. Generally, residents were in disfavor of the attribute level with a fully automated DL model (OR=0.24; *P*<.001), but they preferred attribute levels with lower screening costs and oral explanations by GP. The results for the attributes "organizational form," "missed diagnosis rate," "overdiagnosis rate," and "screening result feedback efficiency" were inconclusive. None of the attribute levels were associated with statistically significant utility differences.

**Table 2.** Characteristics of respondents.

| Respondent and characteristics | Value |
|---|---|
| **Medical staff (n=39)** | |
| Age (years), mean (SD) | 39.67 (6.98) |
| **Institution level, n (%)** | |
| Municipal eye disease control center | 1 (2.56) |
| District-level eye disease control center | 15 (38.46)[a] |
| Community health service center | 23 (58.97) |
| **Position, n (%)** | |
| Institution leader | 7 (17.95) |
| Department leader | 22 (56.41) |
| Eye disease screening mainstay | 10 (25.64) |
| Years in the current position, mean (SD) | 6.73 (5.76) |
| **Resident (n=318)** | |
| Age (years), mean (SD) | 68.62 (6.96) |
| **Sex, n (%)** | |
| Male | 120 (37.74) |
| Female | 198 (62.26) |
| **Education level, n (%)** | |
| Junior high school and below | 216 (67.92) |
| Senior high school | 72 (22.64) |
| Junior college | 21 (6.6) |
| Undergraduate and above | 9 (2.83) |
| **Eye disease, n (%)** | |
| Suspected | 73 (22.96) |
| None | 245 (77.04) |

[a] One respondent from a district-level eye disease control center quit the experiment because of temporary work arrangements. Therefore, although 16 district-level eye disease control centers were included in our study, only 15 key persons from these institutions finished the questionnaire.

**Table 3.** Preferences for using deep learning in community-based eye disease screening.

| Attribute and level | Medical staff[a]<br>OR[b] (95% CI) | Residents<br>OR (95% CI) |
|---|---|---|
| **Diagnostic technology** | | |
| Semiautomated DL[c] model | 2.08 (1.71, 2.52)[d] | 0.89 (0.68, 1.15) |
| Fully automated DL model | 2.39 (1.97, 2.90)[d] | 0.24 (0.20, 0.29)[d] |
| Fully manual diagnosis | Reference | Reference |
| **Organizational form** | | |
| Centralized screening | Reference | Reference |
| Residents’ health self-examination cabin | Not significant | Not significant |
| Opportunity screening in outpatient[e] | Not significant | Not significant |
| **Missed diagnosis rate** | | |
| None | Reference | Reference |
| 5% | Not significant | Not significant |
| 10% | Not significant | Not significant |
| 15% | Not significant | Not significant |
| 20% | Not significant | Not significant |
| **Overdiagnosis rate** | | |
| None | Reference | Reference |
| 5% | 0.88 (0.68, 1.15) | Not significant |
| 10% | 0.61 (0.46, 0.81)[d] | Not significant |
| 15% | 0.63 (0.48, 0.83)[f] | Not significant |
| 20% | 0.51 (0.38, 0.68)[d] | Not significant |
| **Cost[g]** | | |
| Free | Reference | Reference |
| 40 CNY | 0.61 (0.46, 0.83)[f] | 0.75 (0.56, 1.01) |
| 80 CNY | 0.47 (0.35, 0.64)[d] | 0.56 (0.42, 0.74)[d] |
| 120 CNY | 0.39 (0.28, 0.54)[d] | 0.82 (0.51, 1.31) |
| 160 CNY | 0.27 (0.19, 0.38)[d] | 0.78 (0.46, 1.32) |
| 200 CNY | 0.21 (0.15, 0.29)[d] | 0.57 (0.46, 0.71)[d] |
| **Screening result feedback form** | | |
| Screening results | Not significant | 0.52 (0.44, 0.61)[d] |
| Screening results and referral recommendations | Not significant | 0.75 (0.65, 0.87)[d] |
| Screening results, referral recommendations, and oral explanation by GP[h] | Reference | Reference |
| **Screening result feedback efficiency** | | |
| Immediately | Reference | Reference |
| In 2 weeks | 0.68 (0.56, 0.82)[d] | Not significant |
| In 1 month | 0.58 (0.48, 0.70)[d] | Not significant |

[a] In each grid, an OR value over 1 means that the health care services providers were more inclined to this level, while the value less than 1 means that they disliked this level even more.

[b] OR: odds ratio.

[c] DL: deep learning.

[d] *P*<.001.

[e] Residents with chronic diseases and other risk factors would be recommended by general practitioners for eye disease screening during their outpatient follow-up.

[f] *P*=.001.

[g] In 2021, US $1= 6.5 CNY.

[h] GP: general practitioner.

## Discussion

### Principal Findings

To the best of our knowledge, this study is the first to quantitatively estimate both medical staff's and residents' preferences for using DL in community-based eye disease screening in the real world. Since one of the most important questions for achieving universal health coverage in a digital world is whether digital technologies help increase the acceptability of health care services [23,24], our study is significant for the transformation, application, and promotion of this new technology. It was based on the multicenter practices of AI-assisted eye disease screening from 34 medical institutions, where both medical staff and residents under investigation had real service experience of AI. We showed that when compared with a fully manual diagnosis, AI technology was more favored by the medical staff, even after adjusting for the impacts of diagnosis accuracy, cost, and efficiency. However, the residents were in disfavor of the AI technology without doctors' supervision. Furthermore, to meet the medical staff's preference, the accuracy of the AI-assisted eye disease screening technology should be much higher than the FDA's standards. On the contrary, accuracy was not a priority for the residents. They prefer to have the doctors involved in the screening process and leave the choice of accuracy to their general practitioners.

The adoption of DL model for community-based eye disease screening is necessary. Before the development of DL model, the screening relied on ophthalmologists heavily, regardless of conducting traditional face-to-face screening or a telemedicine system [25]. At this stage, continuous eye disease screening was not affordable in most of the countries [6] for two reasons. On the one hand, the limited human resources of the ophthalmologists resulted in extremely high screening costs [5]. On the other hand, the organization of the screening was challenging, requiring the coordination of ophthalmologists, community health centers, and residents at the same time [25]. As a result, in Shanghai, before the adoption of the DL model, each community only could provide screening service to approximately 300 residents per year. On the contrary, after the adoption of DL model, as the ophthalmologist resources were no longer the bottlenecks, the screening use volume dramatically increased to 800 residents per community per year.

Accuracy is regarded as one of the most important considerations in the adoption of DL model. When screening populations with a substantial disease, achieving both high sensitivity and specificity is critical in minimizing both false-positive and false-negative results [26]. The previous studies have shown that it is feasible to meet the mandatory level of accuracy as the primary endpoint with a sensitivity of more than 85% and a specificity of more than 82.5%, which was recommended by the FDA [6,22,27,28]. However, when the DL models were applied in the real world, their accuracy greatly reduced [7]. Therefore, the question is, "what are the medical staff and residents' requirements of the accuracy of AI models in the real world?"

Our study attempted to answer this question from the perspective of medical staff's and residents' preferences in the real-world, community-based eye diseases screening. Although the ideal state is 100% accuracy, under the existing technical conditions, health care service providers must make a trade-off between higher sensitivity and specificity. Both outcomes are important—positive cases should be identified, but this should not come at the cost of overly sensitive screening systems [29].

We showed that if the overdiagnosis rate exceeded 10%, the preferences of the medical staff decreased significantly. Therefore, the specificity of the DL model should be controlled with over 90% accuracy. This does not mean that sensitivity is not important, but rather that the sensitivity standard of the FDA is sufficient. On the one hand, sensitivity is a patient safety criterion, because the primary goal of eye disease screening is to identify the people who are likely to have eye disease and require further evaluation by ophthalmologists [22]. One GP in our study claimed that "the missed diagnosis may harm residents' trust in eye disease screening and reduce their enthusiasm for screening," whereas trust acts as a critical element in medical care [30]. On the other hand, the overdiagnosis rate affects the number of residents who receive an unnecessary referral [22]. A higher overdiagnosis rate means more unnecessary specialist visits, which may lead to unnecessary psychological stress for suspected patients and add further referral costs [31]. Therefore, our results indicated that overdiagnosis would cause resentment from both decision-making and executive agencies.

However, though accuracy is critical for medical staff, results show that the residents do not regard it as a priority. They rather focus on whether the doctors are at the center of medical decision-making [32]. Humans are notoriously poor at comprehending probability and evaluating risk, especially when it pertains to their health or the health of a loved one [33]. In the AI era, although medical knowledge—which forms the basis of decision-making—will be as accessible to the patient as the doctor, most patients need a doctor to understand risk and to communicate this to them [33]. Patients look to the doctors for advice when facing uncertainty in their medical decisions [34]. A study of patient attitudes toward AI use has shown that patients felt their doctors should have the final say in their treatment plans to avoid experiencing the potential harm that might result from mistakes made by health care AI [32].

Therefore, AI tools should be used as decision support tools for human diagnosticians, but not in place of them [35].

When it comes to the other attributes, AI technology with a lower cost and higher feedback efficiency is logically preferable. Cost is an important issue in the adoption of AI-assisted eye disease diagnosis technology. Therefore, it is necessary to conduct health economics evaluation [36]. Fortunately, evidence has shown the cost of screening could be saved by using AI technology, which is mainly attributable to the substantial reduction in human assessment time and workforce without sacrificing screening performance [5].

Traditional ophthalmological diagnosis is heavily dependent on the interpretation of images, which is often subjective and qualitative [37]. Reading these images by trained personnel is neither sustainable nor an efficient use of expertise, and AI technology is essential in facilitating the capture, storage, and interpretation of photographs [17]. From the health system's perspective, the addition of the DL model to fundus photography provides an opportunity to improve this platform for detecting and monitoring retinal diseases on a large scale, and satisfactory results have been obtained [13]. In addition, AI algorithms may bridge the clinical gap [4]. The DL method used for discriminative tasks in ophthalmology, such as diagnosing diabetic retinopathy or age-related macular degeneration, could enhance existing data sets of common and rare ophthalmic diseases without concern for personally identifying information [38]. Other than helping address the limited screening capacity, the DL model may reduce workforce costs and relieve the burden placed on teleophthalmology health care staff [4,39]. The inadequacy of health resources and the vast medical burden may be important reasons for the rapid acceptance the DL method by medical staff.

Regarding feedback efficiency, recent studies have shown that nearly instantaneous feedback may lead to increased patient compliance [5,18]. The most obvious context for the application of AI-assisted diagnosis technology is in primary eye care where the data to be analyzed are complex, the outcomes are simple and well-defined, and the number of people to process is large [18]. In this context, manual diagnosis requires extensive time and energy, whereas AI can work tirelessly and quickly.

### Limitations

The most obvious limitation of our study was that DCE was conducted only in Shanghai. However, as mentioned, Shanghai is one of the pioneers in eye care digital transformation. Therefore, our study is valuable for other regions of the world. The second limitation was that the residents in our experiment were mainly older adults. However, this was consistent with the population that participated in the community-based eye screening in Shanghai because young people mostly participated in physical examinations at their workplace.

### Conclusion

In conclusion, to meet the actual preferences of medical staff and residents for using AI in the community-based eye disease screening, the DL model under doctors' supervision is strongly recommended, and the specificity of the model should be more than 90%, which is higher than the FDA standard. In addition, digital transformation should help medical staff move away from heavy and repetitive work; however, it should not reduce their involvement in the health care service. Instead, medical staff should spend more time on communicating with residents.

## Acknowledgments

This study was funded by the Shanghai Public Health Three-Year Action Plan (No. GWV10.1-XK6), Science and Technology Commission of Shanghai Municipality (No. 20DZ1100200), Shanghai Hospital Development Center (SHDC12021613), Shanghai Eye Disease Control and Treatment Center (20LC01002), and Shanghai Municipal Health Commission (2022HP61).

We gratefully acknowledge the following investigators who also made contributions to the study: Yajun Peng, Tao Yu, Yao Yin, and Dan Qian.

## Availability of Data and Materials

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

## Conflicts of Interest

None declared.

## References

1. GBD 2019 BlindnessVision Impairment Collaborators, Vision Loss Expert Group of the Global Burden of Disease Study. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. Lancet Glob Health 2021 Feb;9(2):e130-e143 [FREE Full text] [doi: 10.1016/S2214-109X(20)30425-3] [Medline: 33275950]
2. Eye care competency framework. World Health Organization. URL: https://www.who.int/publications/i/item/9789240048416 [accessed 2022-09-09]

3. Cheng C, Wang N, Wong TY, Congdon N, He M, Wang YX, Vision Loss Expert Group of the Global Burden of Disease Study. Prevalence and causes of vision loss in East Asia in 2015: magnitude, temporal trends and projections. Br J Ophthalmol 2020 May 28;104(5):616-622. [doi: 10.1136/bjophthalmol-2018-313308] [Medline: 31462416]
4. Ting DS, Lee AY, Wong TY. An ophthalmologist's guide to deciphering studies in artificial intelligence. Ophthalmology 2019 Nov;126(11):1475-1479 [FREE Full text] [doi: 10.1016/j.ophtha.2019.09.014] [Medline: 31635697]
5. Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekeran DV, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. The Lancet Digital Health 2020 May;2(5):e240-e249. [doi: 10.1016/s2589-7500(20)30060-1]
6. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. Prog Retin Eye Res 2018 Nov;67:1-29 [FREE Full text] [doi: 10.1016/j.preteyeres.2018.07.004] [Medline: 30076935]
7. Lee A, Yanagihara R, Lee C, Blazes M, Jung H, Chee Y, et al. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. Diabetes Care 2021 May;44(5):1168-1175 [FREE Full text] [doi: 10.2337/dc20-1877] [Medline: 33402366]
8. Ryan M. Discrete choice experiments in health care. BMJ 2004 Feb 14;328(7436):360-361 [FREE Full text] [doi: 10.1136/bmj.328.7436.360] [Medline: 14962852]
9. de Bekker-Grob EW, Swait JD, Kassahun HT, Bliemer MC, Jonker MF, Veldwijk J, et al. Are healthcare choices predictable? The impact of discrete choice experiment designs and models. Value Health 2019 Sep;22(9):1050-1062 [FREE Full text] [doi: 10.1016/j.jval.2019.04.1924] [Medline: 31511182]
10. Xu Y, Wang Y, Liu B, Tang L, Lv L, Ke X, et al. The diagnostic accuracy of an intelligent and automated fundus disease image assessment system with lesion quantitative function (SmartEye) in diabetic patients. BMC Ophthalmol 2019 Aug 14;19(1):184 [FREE Full text] [doi: 10.1186/s12886-019-1196-9] [Medline: 31412800]
11. Dai L, Wu L, Li H, Cai C, Wu Q, Kong H, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. Nat Commun 2021 May 28;12(1):3242 [FREE Full text] [doi: 10.1038/s41467-021-23458-5] [Medline: 34050158]
12. Li F, Wang Y, Xu T, Dong L, Yan L, Jiang M, et al. Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. Eye (Lond) 2022 Jul 01;36(7):1433-1441. [doi: 10.1038/s41433-021-01552-8] [Medline: 34211137]
13. Lin D, Xiong J, Liu C, Zhao L, Li Z, Yu S, et al. Application of Comprehensive Artificial intelligence Retinal Expert (CARE) system: a national real-world evidence study. The Lancet Digital Health 2021 Aug;3(8):e486-e495. [doi: 10.1016/s2589-7500(21)00086-8]
14. Cen L, Ji J, Lin J, Ju S, Lin H, Li T, et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. Nat Commun 2021 Aug 10;12(1):4828 [FREE Full text] [doi: 10.1038/s41467-021-25138-w] [Medline: 34376678]
15. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16(1):3-13 [FREE Full text] [doi: 10.1016/j.jval.2012.08.2223] [Medline: 23337210]
16. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
17. Li JO, Liu H, Ting DS, Jeon S, Chan RP, Kim JE, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. Prog Retin Eye Res 2021 May;82:100900 [FREE Full text] [doi: 10.1016/j.preteyeres.2020.100900] [Medline: 32898686]
18. Lee A, Taylor P, Kalpathy-Cramer J, Tufail A. Machine learning has arrived!. Ophthalmology 2017 Dec;124(12):1726-1728. [doi: 10.1016/j.ophtha.2017.08.046] [Medline: 29157423]
19. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 2015 Feb;144(1):114-126. [doi: 10.1037/xge0000033] [Medline: 25401381]
20. Kuhfeld WF. Marketing research methods in SAS: experimental design, choice, conjoint, and graphical techniques. SAS Institute Inc. URL: https://support.sas.com/techsup/technote/mr2010.pdf [accessed 2022-09-09]
21. Sculpher M, Bryan S, Fry P, de Winter P, Payne H, Emberton M. Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. BMJ 2004 Feb 14;328(7436):382 [FREE Full text] [doi: 10.1136/bmj.37972.497234.44] [Medline: 14751919]
22. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018 Aug 28;1(1):39 [FREE Full text] [doi: 10.1038/s41746-018-0040-6] [Medline: 31304320]
23. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, et al. The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world. The Lancet 2021 Nov;398(10312):1727-1776. [doi: 10.1016/s0140-6736(21)01824-9]
24. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. The Lancet 2020 May;395(10236):1579-1586. [doi: 10.1016/s0140-6736(20)30226-9]
25. Li R, Yang Z, Zhang Y, Bai W, Du Y, Sun R, et al. Cost-effectiveness and cost-utility of traditional and telemedicine combined population-based age-related macular degeneration and diabetic retinopathy screening in rural and urban China.

Lancet Reg Health West Pac 2022 Jun;23:100435 [FREE Full text] [doi: 10.1016/j.lanwpc.2022.100435] [Medline: 35355615]

26. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
27. Bhaskaranand M, Ramachandra C, Bhat S, Cuadros J, Nittala MG, Sadda SR, et al. The value of automated diabetic retinopathy screening with the EyeArt system: A study of more than 100,000 consecutive encounters from people with diabetes. Diabetes Technol Ther 2019 Nov;21(11):635-643 [FREE Full text] [doi: 10.1089/dia.2019.0164] [Medline: 31335200]
28. Ming S, Xie K, Lei X, Yang Y, Zhao Z, Li S, et al. Evaluation of a novel artificial intelligence-based screening system for diabetic retinopathy in community of China: a real-world study. Int Ophthalmol 2021 Apr 03;41(4):1291-1299. [doi: 10.1007/s10792-020-01685-x] [Medline: 33389425]
29. Tufail A, Kapetanakis VV, Salas-Vega S, Egan C, Rudisill C, Owen CG, et al. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. Health Technol Assess 2016 Dec;20(92):1-72 [FREE Full text] [doi: 10.3310/hta20920] [Medline: 27981917]
30. Owsley C, McGwin G, Scilley K, Girkin CA, Phillips JM, Searcey K. Perceived barriers to care and attitudes about vision and eye care: focus groups with older African Americans and eye care providers. Invest Ophthalmol Vis Sci 2006 Jul 01;47(7):2797-2802. [doi: 10.1167/iovs.06-0107] [Medline: 16799016]
31. Wong RL, Tsang C, Wong DS, McGhee S, Lam C, Lian J, et al. Are we making good use of our public resources? The false-positive rate of screening by fundus photography for diabetic macular oedema. Hong Kong Med J 2017 Aug 7;23(4):356-364 [FREE Full text] [doi: 10.12809/hkmj166078] [Medline: 28684650]
32. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med 2021 Sep 21;4(1):140 [FREE Full text] [doi: 10.1038/s41746-021-00509-1] [Medline: 34548621]
33. Liu X, Keane PA, Denniston AK. Time to regenerate: the doctor in the age of artificial intelligence. J R Soc Med 2018 Apr 12;111(4):113-116 [FREE Full text] [doi: 10.1177/0141076818762648] [Medline: 29648509]
34. Dietvorst BJ, Bharti S. People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. Psychol Sci 2020 Oct 11;31(10):1302-1314. [doi: 10.1177/0956797620948841] [Medline: 32916083]
35. Sarwar S, Dent A, Faust K, Richer M, Djuric U, Van Ommeren R, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. NPJ Digit Med 2019 Apr 26;2(1):28 [FREE Full text] [doi: 10.1038/s41746-019-0106-0] [Medline: 31304375]
36. Ruamviboonsuk P, Chantra S, Seresirikachorn K, Ruamviboonsuk V, Sangroongruangsri S. Economic evaluations of artificial intelligence in ophthalmology. Asia Pac J Ophthalmol (Phila) 2021 Jul 13;10(3):307-316. [doi: 10.1097/APO.0000000000000403] [Medline: 34261102]
37. Coyner AS, Campbell JP, Chiang MF. Demystifying the jargon: the bridge between ophthalmology and artificial intelligence. Ophthalmol Retina 2019 Apr;3(4):291-293 [FREE Full text] [doi: 10.1016/j.oret.2018.12.008] [Medline: 31014678]
38. Burlina PM, Joshi N, Pacheco KD, Liu TYA, Bressler NM. Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. JAMA Ophthalmol 2019 Mar 01;137(3):258-264 [FREE Full text] [doi: 10.1001/jamaophthalmol.2018.6156] [Medline: 30629091]
39. Domalpally A, Channa R. Real-world validation of artificial intelligence algorithms for ophthalmic imaging. Lancet Digit Health 2021 Aug;3(8):e463-e464 [FREE Full text] [doi: 10.1016/S2589-7500(21)00140-0] [Medline: 34325850]

## Abbreviations

**AI:** artificial intelligence
**DCE:** discrete choice experiment
**DL:** deep learning
**FDA:** Food and Drug Administration
**GP:** general practitioner
**OR:** odds ratio

*Edited by R Kukafka; submitted 12.06.22; peer-reviewed by Z Hou, W Li; comments to author 13.07.22; revised version received 08.08.22; accepted 02.09.22; published 20.09.22.*

*Please cite as:*
*Lin S, Li L, Zou H, Xu Y, Lu L*
*Medical Staff and Resident Preferences for Using Deep Learning in Eye Disease Screening: Discrete Choice Experiment*
*J Med Internet Res 2022;24(9):e40249*
*URL: https://www.jmir.org/2022/9/e40249*
*doi:10.2196/40249*
*PMID:36125854*

Original Paper

# Digital Health Literacy as a Predictor of Awareness, Engagement, and Use of a National Web-Based Personal Health Record: Population-Based Survey Study

Christina Cheng[1], PhD; Emma Gearon[2], PhD; Melanie Hawkins[1], PhD; Crystal McPhee[3], MPH; Lisa Hanna[4], PhD; Roy Batterham[5], MEd; Richard H Osborne[1], PhD

[1]Centre for Global Health and Equity, School of Health Sciences, Swinburne University of Technology, Hawthorn, Australia
[2]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia
[3]Beyondblue, Melbourne, Australia
[4]School of Health and Social Development, Faculty of Health, Deakin University, Geelong, Australia
[5]Faculty of Public Health, Thammasat University, Bangkok, Thailand

**Corresponding Author:**
Christina Cheng, PhD
Centre for Global Health and Equity
School of Health Sciences
Swinburne University of Technology
Room 907, Level 9, AMDC Building
453/469-477 Burwood Road
Hawthorn, Victoria 3122
Australia
Phone: 61 392145470
Email: cccheng@swin.edu.au

## Abstract

**Background:** Web-based personal health records (PHRs) have the potential to improve the quality, accuracy, and timeliness of health care. However, the international uptake of web-based PHRs has been slow. Populations experiencing disadvantages are less likely to use web-based PHRs, potentially widening health inequities within and among countries.

**Objective:** With limited understanding of the predictors of community uptake and use of web-based PHR, the aim of this study was to identify the predictors of awareness, engagement, and use of the Australian national web-based PHR, My Health Record (MyHR).

**Methods:** A population-based survey of adult participants residing in regional Victoria, Australia, was conducted in 2018 using telephone interviews. Logistic regression, adjusted for age, was used to assess the relationship among digital health literacy, health literacy, and demographic characteristics, and the 3 dependent variables of MyHR: awareness, engagement, and use. Digital health literacy and health literacy were measured using multidimensional tools, using all 7 scales of the eHealth Literacy Questionnaire and 4 out of the 9 scales of the Health Literacy Questionnaire.

**Results:** A total of 998 responses were analyzed. Many elements of digital health literacy were strongly associated with MyHR awareness, engagement, and use. A 1-unit increase in each of the 7 eHealth Literacy Questionnaire scales was associated with a 2- to 4-fold increase in the odds of using MyHR: *using technology to process health information* (odds ratio [OR] 4.14, 95% CI 2.34-7.31), *understanding of health concepts and language* (OR 2.25, 95% CI 1.08-4.69), *ability to actively engage with digital services* (OR 4.44, 95% CI 2.55-7.75), *feel safe and in control* (OR 2.36, 95% CI 1.43-3.88), *motivated to engage with digital services* (OR 4.24, 95% CI 2.36-7.61), *access to digital services that work* (OR 2.49, 95% CI 1.32-4.69), and *digital services that suit individual needs* (OR 3.48, 95% CI 1.97-6.15). The Health Literacy Questionnaire scales of health care support, actively managing health, and social support were also associated with a 1- to 2-fold increase in the odds of using MyHR. Using the internet to search for health information was another strong predictor; however, older people and those with less education were less likely to use MyHR.

**Conclusions:** This study revealed strong and consistent patterns of association between digital health literacy and the use of a web-based PHR. The results indicate potential actions for promoting PHR uptake, including improving digital technology and skill experiences that may improve digital health literacy and willingness to engage in web-based PHR. Uptake may also be improved through more responsive digital services, strengthened health care, and better social support. A holistic approach, including targeted solutions, is needed to ensure that web-based PHR can realize its full potential to help reduce health inequities.

***(J Med Internet Res 2022;24(9):e35772)*** doi:10.2196/35772

**KEYWORDS**

eHealth; mobile health; mHealth; health literacy; health equity; electronic health records; vulnerable populations; disadvantaged populations

# Introduction

## Background

Digital technologies have enabled the storage of personal health information in a web-based environment where people can keep track of and access their health records as needed [1-4]. The World Health Organization considers a national web-based personal health record (PHR) an important component of universal health coverage, given its potential to improve the quality, accuracy, and timeliness of health care [5]. In 2016, half (47%) of the member states responding to a survey indicated that they had introduced a national web-based PHR system [5,6].

In 2012, Australia, a country with universal health coverage, rolled out a national web-based PHR system, My Health Record (MyHR), aiming to provide Australians with *safer, faster and more efficient* health care [7,8]. MyHR is a secure web-based summary of a person's health information, allowing people to control and share personal health information with their health care providers anytime and anywhere, thereby improving communication between clinicians and patients. The system is available in more than 18 languages and is enabled for people with low vision and blindness to ensure access to people with various needs [9]. However, MyHR uptake was slow. Only approximately 21% (5.2 million out of 24.6 million) of Australia's total population signed up for MyHR in 2017, after 5 years of rollout [10]. In 2018, the Australian Government announced that MyHR would become an opt-out system, meaning that people would automatically be enrolled in MyHR unless they elected not to be during the opt-out period between July 2018 and January 2019 [11].

The slow uptake of web-based PHRs is not unique to Australia [1]. A study in the United States reported that only 20% of survey participants used web-based PHRs [12]. After an investment of £8 million (approximately US $11 million) in a web-based PHR in the United Kingdom in 2007, the system HealthSpace was abandoned in 2011 owing to low adoption. The postevaluation report commented that the lack of health literacy and digital literacy in some users might be one of the reasons for nonadoption [13].

## Digital Health Literacy

Digital health literacy (also called eHealth literacy) can be one of the deciding factors when determining whether to use a digital health system [14-16]. Digital health literacy refers to an individual's "ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [14]. This concept has continued to evolve with the ever-changing digital landscape since it was first conceived in 2006. Nevertheless, it is grounded in health literacy, which is defined as "people's knowledge, confidence, and comfort, which accumulate through daily activities, social interactions, and across generations, to access, understand, appraise, remember, and use information about health and health care for the health and wellbeing of themselves and those around them" (World Health Organization, unpublished data, September 2022).

PHRs have the potential to reduce health inequities [17,18]. However, this potential may be derailed by the digital divide [2,3], which refers to inequitable access, use, and outcomes of technology use among subgroups of society, because people with higher income or education are more likely to have better access or skills to use technology than other groups, such as culturally diverse minorities, rural residents, or people with lower income or education [19-22]. These groups are also less likely to use web-based PHRs [3,23] and face more digital health literacy challenges [24-27]. Hence, well-intentioned efforts to reduce health inequities must be implemented with care to ensure that they do not worsen health disparities [2,3,23]. Understanding the predictors of using web-based PHRs is an essential first step to avoid this pitfall.

To date, research on the predictors of web-based PHR uptake is limited. Only a few studies could be identified, and they mostly focused on age, education, health status, computer skills, and experiences in web-based information seeking [28-33]. Only one study has specifically examined digital health literacy [33]. Given the limited understanding of the predictors of uptake and use of web-based PHR, the aim of this study was to determine the predictors of awareness, engagement, and use of the Australian personal web-based health record MyHR during the MyHR opt-out period, with a focus on digital health literacy. Regarding predictors, this study referred to the statistical procedure used to identify factors associated with PHR uptake and use, not as an indication of the causal relationship with these factors.

# Methods

## Study Design

This study was part of a larger study conducted in the city of Ballarat and the surrounding regional area in Victoria, Australia [34]. Victoria is the second most populous state in Australia, with a population of approximately 6.7 million as of June 2021 [35]. The city of Ballarat is situated in the regional area of north-central Victoria. This area was selected as it was a trial site for the early implementation of the MyHR [36] and chosen in partnership with the Australian Digital Health Agency, the commonwealth entity established to oversee the operation and evolution of Australian digital health capability [37]. The aim of the larger study was to generate insights into how to maximize the uptake and use of MyHR and other digital technologies. The study was an application of the Optimizing Health Literacy and Access (Ophelia) process, which involved identifying local needs and then using a co-design approach to engage stakeholders (consumers, practitioners, and managers) to generate fit-for-purpose solutions through insights from local wisdom [38]. This paper presents the findings from the needs assessment phase of the study.

### Data Collection

A computer-assisted telephone interview (CATI) survey was administered from October 1, 2018, to October 31, 2018, to participants from the general population. Participants were eligible for inclusion if they were aged >18 years, able to complete a telephone survey in English, and resided in the Ballarat Goldfields region of Victoria, Australia. There were no quotas for age or sex. The interviews were conducted by a contract research company (Strahan Research Pty Ltd). The interviewer team underwent specialized training with the project team. A pilot phase was conducted, with the responses reviewed before the formal implementation of the survey. Each interview started with a description of the project requirements and the consent process, followed by the survey. Only respondents who consented to participate were included in this study. Each interview lasted approximately 17 minutes. The completed CATI surveys were deidentified before being provided to the project team.

To ensure the sample was stratified by socioeconomic position, it was drawn using systematic random sampling: all postcodes of the region were ordered by an area-level marker of socioeconomic position based on the Index of Relative Socioeconomic Disadvantage, an index that summarizes the economic and social conditions of people within an area [39]. A database of both landline telephone and mobile phone numbers was matched to the postal areas, and the sample was drawn using a random start fixed interval sampling technique. The fixed interval was calculated by dividing the total population of the Ballarat Goldfields region by the desired sample size of 1000.

### Ethics Approval

Ethics approval was obtained from the Deakin University Human Research Ethics Committee (HEAG-H 157_2018).

### Survey Instruments

To examine the possible predictors of awareness, engagement, and use of MyHR, digital health literacy assessment, demographic data, and use of health services data were collected. Digital health literacy was assessed using the eHealth Literacy Questionnaire (eHLQ) [40], complemented by the Health Literacy Questionnaire (HLQ) [41] to provide context for the eHLQ results. The CATI survey consisted of all 7 scales of the eHLQ and 4 scales from the HLQ, followed by questions about the participants' demographics, health status, use of physical and digital health services, and experience of MyHR.

The eHLQ and HLQ were developed using a grounded validity-driven approach [42] to assess the multidimensional concepts of digital health literacy and health literacy. Both tools, at construction and initial validity testing, were found to be psychometrically robust [40,41], with later studies presenting acceptable to strong psychometric properties when the tools were used in different contexts [43-51]. The eHLQ consists of 35 items, with 7 scales representing the 7 dimensions of digital health literacy (see Multimedia Appendix 1 for scale definitions):

1. Using technology to process health information
2. Understanding of health concepts and language
3. Ability to actively engage with digital services
4. Feel safe and in control
5. Motivated to engage with digital services
6. Access to digital services that work
7. Digital services that suit individual needs

A 4-point response option of strongly disagree, disagree, agree, and strongly agree was used in the eHLQ. Scale scores were calculated by averaging the item scores within each scale with equal weighting, yielding 7 scale scores, each with a score range of 1 to 4 [40].

The full HLQ consists of 44 items across 9 scales [41] (see Multimedia Appendix 1 for scale definitions). However, to reduce the length of the telephone interviews, only 4 scales were used in this survey (in italics as follows):

1. *Feeling understood and supported by health care providers*
2. Having sufficient information to manage my health
3. *Actively managing my health*
4. *Social support for health*
5. Appraisal of health information
6. Ability to actively engage with health care providers
7. *Navigating the health care system*
8. Ability to find good information
9. Understand health information well enough to know what to do

Scales 1 and 4 of the HLQ reflect a social orientation that is positive in managing one's health, and a PHR is a potentially valuable tool for this, and scale 1 also reflects the quality of communication and trust in health care providers, which is regarded as potentially important as health care providers are likely to be a key source of information for people about whether they use the MyHR. Scales 3 and 7 reflect general engagement with health and health care and are expected to provide context regarding a person's level of interest in MyHR.

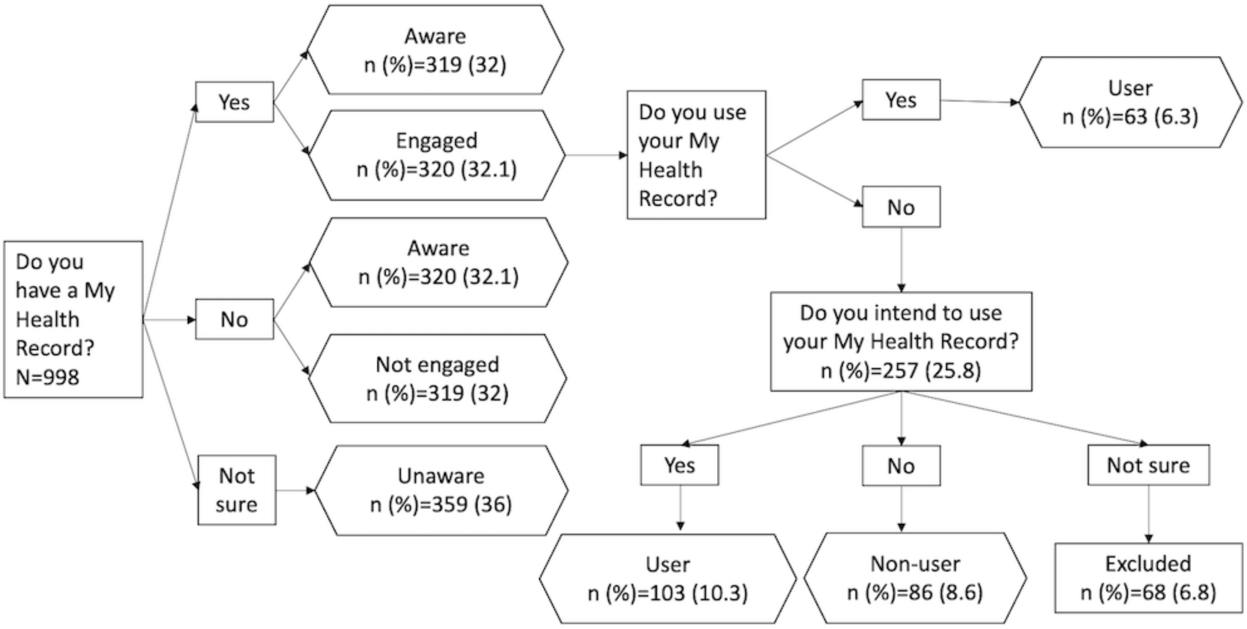
The HLQ was subjected to rigorous validity testing in the initial validation study. The unidimensionality of each scale was established with evidence of satisfactory fit for each of the nine 1-factor models, and composite reliability of each scale ranged from 0.77 to 0.90 [41]. Validity evidence for the English version was further confirmed in subsequent validation studies in Australia [43,47,52,53], the English version in other contexts [54] and elsewhere [44-46,55-57], confirming that the HLQ scales can be used independently to measure different dimensions of health literacy.

Similar to the eHLQ, scales 1, 3, and 4 of the HLQ also use the 4-point response options of strongly disagree to strongly agree. For items of scale 7, a 5-point response option of cannot do or always difficult, usually difficult, sometimes difficult, usually easy, and always easy is used. The calculation of scale scores is the same as that of the eHLQ, except for HLQ scale 7, which has a score range of 1 to 5.

For the assessment of demographic background, age was measured as a continuous variable (years); sex was measured as female or male; highest educational attainment was measured in 4 categories (did not complete secondary school; completed secondary school; trade, apprenticeship, certificate, or diploma; and university); number of chronic health conditions was a discrete variable, calculated as the sum of any of the 8 conditions—asthma, arthritis, anxiety, cancer, cardiovascular disease or heart problems, chronic pain, depression, and diabetes (no condition; 1 condition; and 2 or more conditions); self-rated health was measured on a discrete scale with 6 options, ranging from excellent to very poor; number of contacts with a health professional in the past 12 months was measured in 4 categories (≥13 times; 7-12 times; 2-6 times; and 0-1 time); and use of the internet to search for health-related information in the past 12 months was measured as a binary variable (yes or no).

The 3 variables—MyHR awareness, MyHR engagement, and MyHR use or intention to use—were determined by 3 questions in the survey. MyHR awareness was measured in the total sample using the question *Do you have a My Health Record?* (yes, no, or not sure). Participants who answered not sure were classified as *unaware*, and those who responded either yes or no were classified as *aware*, because it was a clear indication that they were aware of the existence of MyHR. For MyHR engagement, participants who answered yes to the aforementioned question were classified as *engaged*, and those who answered no, as *not engaged*. MyHR use was measured in the *engaged* subpopulation, who were asked, *Do you use your My Health Record?* (yes and no). Participants who answered yes were categorized as *user*, whereas participants who responded with a no were directed to the question, *Do you intend to use My Health Record?* (yes, no, or not sure). Participants who responded yes to this final question were classified as *intend to use* and were categorized as *user*. Those who responded that they did not intend to use MyHR were categorized as *nonuser*. To ensure that the variable gave a clear signal, *engaged* participants who chose the *not sure* option were excluded from analysis. Figure 1 shows the flow diagram for identifying awareness (aware or unaware), engagement (engaged or not engaged), and use or intention to use (user or nonuser).

**Figure 1.** Flow diagram for identifying My Health Record awareness (aware or not aware), engagement (engaged or not engaged), and use or intention to use (user or nonuser; N=998).



## Statistical Analysis

As this was an exploratory study, a sample size of 1000 was chosen to provide ample power to detect small to moderate differences across target sex, age, education, and other subgroups (ie, 0.1-0.2–unit differences among subgroups on the 4-point response range on the eHLQ scales, with a power of at least 0.8, α of .05, and SD of 0.7), conservatively based on sample characteristics reported in the validity testing paper by Kayser et al [40].

The data were analyzed using Stata (version 15; StataCorp) [58]. Descriptive statistics of demographics and mean and 95% CIs were calculated for eHLQ and HLQ scores. The demographics of the sample were compared with those of the population of the Ballarat Goldfields region of the Western Victoria Primary Health Network (WVPHN) [59] to determine the representativeness of the sample. Primary health networks are independent primary health care organizations established throughout Australia by the Australian Government in 2015 to reform the primary health care system using a patient-oriented approach to medical services in their regions. The Ballarat

Goldfields region of Victoria is part of the WVPHN, covering 21 areas in the western district of Victoria, Australia.

Logistic regression, adjusted for age, was performed to assess the relationship among independent variables, including demographic characteristics, digital health literacy, and health literacy, and the 3 dependent variables of MyHR: awareness, engagement, and use. Data collected via the eHLQ and HLQ were treated as continuous variables. The study was explorative, and no hypothesis was established, with independent variables for analysis selected based on factors plausibly associated with engagement with digital health technologies.

To ensure that the assumptions of the logistic regression were met, independent variables were collapsed as required to ensure that there were 20 or more people in each stratum. As such, the self-rated health variable was collapsed into 5 categories for all 3 dependent variables, with poor and very poor combined. The number of contacts with a health professional was collapsed into 3 categories for MyHR engagement and 2 categories for MyHR use. Education was collapsed into 3 categories, and self-rated health was 2 categories for MyHR use.

## Results

### Sample Characteristics

A total of 2839 calls were made and 1000 responses were recorded, resulting in a response rate of 35.2%. Of the 1000 CATI survey responses, 2 participants were found to be ineligible for the survey because of age, leaving the data to 998 participants for analysis. For these 998 participants, there were no missing data for any of the variables. Table 1 presents the descriptive statistics of the final sample. Compared with the Ballarat Goldfields region population, the sample had a similar proportion of people who were identified as Aboriginal or Torres Strait Islander (18/998, 1.8% for the survey compared with 2204/157,472, 1.4%) and a similar proportion of women (536/998, 53.7% for the survey compared with 80,594/157,472, 51.18%). However, the survey sample had a higher proportion of people aged 55 to 70 years (384/998, 38.5% for the survey vs 31,344/157,472, 19.9%) and >70 years (441/998, 41.2% for the survey vs 20,091/157,472, 12.76%), and a higher proportion of people with a university education (307/998, 30.8% for the survey vs 26,770/157,472, 17%; see Multimedia Appendix 2).

Approximately half of the sample (462/998, 46.3%) were male participants, and 79.7% (796/998) were aged >55 years. Approximately one-fourth of the sample (272/998, 27.3%) were living with 1 chronic condition, and 14.1% (141/998) reported having 2 or more of the 8 chronic conditions presented, with arthritis (144/998, 14.4%) and cardiovascular disease or heart problems (129/998, 12.9%) being the most commonly reported. More than half of the sample (578/998, 57.9%) reported having used the internet in the past year to search for health-related information (Table 1). Approximately two-thirds of the participants (639/998, 64%) were aware of MyHR, and 32.1% (320/998) were engaged with MyHR. However, only 6.3% (63/998) were current users of MyHR, and 10.3% (103/998) of the participants indicated that they intended to use the system (Figure 1).

For digital health literacy, participants generally disagreed (for mean score of <2.5 on a scale, ranging from 1=strongly disagree to 4=strongly agree) that they were using technology to process health information (eHLQ scale 1), had the ability to engage with digital services (eHLQ scale 3), were motivated to engage with digital services (eHLQ scale 5), and had digital services that suit their needs (eHLQ scale 7). However, they reported a generally good understanding of health concepts (eHLQ scale 2). For health literacy, participants agreed that they had good social (HLQ scale 4) and health care support (HLQ scale 1) and were actively managing their health (HLQ scale 3; Table 1).

**Table 1.** Participant characteristics (N=998).

| Characteristics | Value |
|---|---|
| **Sex, n (%)** | |
| Female | 536 (53.7) |
| Male | 462 (46.3) |
| **Age (years), n (%)** | |
| 18 to <35 | 33 (3.3) |
| 35 to <55 | 169 (16.9) |
| 55 to <75 | 562 (56.3) |
| ≥75 | 234 (23.4) |
| Spoke English at home, n (%) | 8 (0.8) |
| Aboriginal or Torres Strait Islander, n (%) | 18 (1.8) |
| **Education level, n (%)** | |
| Completed primary school or below | 91 (9.1) |
| Did not complete secondary school | 218 (21.8) |
| Completed secondary school | 136 (13.6) |
| Trade, apprenticeship, certificate, or diploma | 246 (24.6) |
| University | 307 (30.8) |
| **Reported long-standing conditions[a], n (%)** | |
| Arthritis | 144 (14.4) |
| Asthma | 42 (4.2) |
| Cancer | 53 (5.3) |
| Cardiovascular disease or heart problems | 129 (12.9) |
| Diabetes | 84 (8.4) |
| Anxiety | 29 (2.9) |
| Depression | 45 (4.5) |
| Chronic pain | 87 (8.7) |
| **Number of long-standing conditions reported[b], n (%)** | |
| 0 | 585 (58.6) |
| 1 | 272 (27.3) |
| 2 | 102 (10.2) |
| 3 or more | 39 (3.9) |
| **Number of contacts with a health professional in the past 12 months, n (%)** | |
| 12 or more | 269 (27) |
| 7 to 11 | 196 (19.6) |
| 2 to 6 | 446 (44.7) |
| Once | 61 (6.1) |
| Not at all | 26 (2.6) |
| **Self-rated health, n (%)** | |
| Excellent | 170 (17) |
| Very good | 270 (27.1) |
| Good | 293 (29.4) |
| Fair | 165 (16.5) |

| Characteristics | Value |
|---|---|
| Poor | 76 (7.6) |
| Very poor | 24 (2.4) |
| **Whether internet is used to search for health-related information, n (%)** | |
| Yes | 578 (57.9) |
| No | 420 (42.1) |
| **eHealth Literacy Questionnaire scales (score range 1 to 4), mean (95% CI)** | |
| 1. Using technology to process health information | 2.34 (2.31-2.38) |
| 2. Understanding of health concepts and language | 2.93 (2.91-2.96) |
| 3. Ability to actively engage with digital services | 2.46 (2.42-2.49) |
| 4. Feel safe and in control | 2.59 (2.56-2.63) |
| 5. Motivated to engage with digital services | 2.40 (2.37-2.44) |
| 6. Access to digital services that work | 2.51 (2.48-2.54) |
| 7. Digital services that suit individual needs | 2.39 (2.35-2.42) |
| **Health Literacy Questionnaire scales (score range 1 to 4), mean (95% CI)** | |
| 1. Feeling understood and supported by health care providers | 3.20 (3.10-3.20) |
| 3. Actively managing my health | 3.02 (2.99-3.05) |
| 4. Social support for health | 3.05 (3.02-3.08) |
| **Health Literacy Questionnaire (score range 1 to 5), mean (95% CI) scale** | |
| 7. Navigating the health care system | 3.95 (3.90-3.99) |

[a] Participants might select more than one condition.

[b] Number based on long-standing conditions selected by participants (range 0-8).

## Predictors of MyHR Awareness, Engagement, and Use or Intention to Use

### *Awareness*

Among all independent variables tested, digital health literacy was a strong predictor of MyHR awareness (Table 2 and Multimedia Appendix 3). With the exception of the eHLQ scale 2, *Understanding of health concepts and language*, a 1-unit increase in each of the other 6 eHLQ scales was associated with a 1.28- to 1.99-fold increase in the odds of MyHR awareness, with scale 6, *Access to digital services that work*, being the strongest. However, the 4 HLQ scales, representing health literacy, were not predictors of MyHR awareness.

For demographic factors, only sex and internet use to search for health-related information in the past 12 months were moderately strong predictors of MyHR awareness. Being female (odds ratio [OR] 1.44, 95% CI 1.11-1.87) and using the internet for health information (OR 1.52, 95% CI 1.15-2.01) were associated with higher odds of MyHR awareness compared with being male and not using the internet (Multimedia Appendix 3).

**Table 2.** Logistic regression predicting likelihood of My Health Record awareness, engagement, and use or intention to use, as measured by the eHealth Literacy Questionnaire and Health Literacy Questionnaire Scales.

| Scale used | Awareness (aware n=639; unaware n=359) | | Engagement (engaged n=320; not engaged n=319) | | Use or intention to use (user n=166; nonuser n=86) | |
|---|---|---|---|---|---|---|
| | Odds ratio (95% CI) | *P* value | Odds ratio (95% CI) | *P* value | Odds ratio (95% CI) | *P* value |
| **eHealth Literacy Questionnaire scales** | | | | | | |
| 1. Using technology to process health information | 1.77 (1.42-2.22) | <.001 | 1.81 (1.35-2.42) | .001 | 4.14 (2.34-7.31) | <.001 |
| 2. Understanding of health concepts and language | 1.28 (0.92-1.77) | .14 | 2.62 (1.70-4.03) | <.001 | 2.25 (1.08-4.69) | .03 |
| 3. Ability to actively engage with digital services | 1.53 (1.25-1.89) | <.001 | 2.12 (1.60-2.81) | <.001 | 4.44 (2.55-7.75) | <.001 |
| 4. Feel safe and in control | 1.47 (1.17-1.85) | .001 | 1.61 (1.20-2.14) | .001 | 2.36 (1.43-3.88) | .001 |
| 5. Motivated to engage with digital services | 1.75 (1.40-2.19) | <.001 | 2.00 (1.48-2.71) | <.001 | 4.24 (2.36-7.61) | <.001 |
| 6. Access to digital services that work | 1.99 (1.51-2.64) | <.001 | 1.90 (1.33-2.70) | <.001 | 2.49 (1.32-4.69) | .005 |
| 7. Digital services that suit individual needs | 1.63 (1.30-2.04) | <.001 | 1.89 (1.40-2.55) | <.001 | 3.48 (1.97-6.15) | <.001 |
| **Health Literacy Questionnaire scales** | | | | | | |
| 1. Feeling understood and supported by health care providers | 1.15 (0.90-1.48) | .26 | 1.63 (1.19-2.22) | .002 | 1.89 (1.10-3.27) | .02 |
| 3. Actively managing my health | 0.89 (0.66-1.21) | .46 | 1.23 (0.87-1.80) | .24 | 2.28 (1.18-4.38) | .01 |
| 4. Social support for health | 1.02 (0.78-1.33) | .89 | 1.74 (1.25-2.42) | .001 | 2.10 (1.15-3.84) | .02 |
| 7. Navigating the health care system | 1.10 (0.92-1.31) | .29 | 1.15 (0.93-1.42) | .20 | 1.24 (0.87-1.75) | .23 |

### *Engagement*

All 7 dimensions of digital health literacy, as assessed by the eHLQ, were strongly associated with MyHR engagement (Table 2 and Multimedia Appendix 4) compared with the other independent variables. A 1-unit increase on each scale was associated with a 1.61- to 2.62-fold increase in the odds of MyHR engagement. Unlike MyHR awareness, in which scale 2, *Understanding of health concepts and language*, was not a predictor; this dimension demonstrated the strongest association with engagement among all the digital health literacy dimensions.

Although the 4 health literacy dimensions assessed by the HLQ were not predictors of awareness, HLQ scale 1, *Feeling understood and supported by health care providers*, and scale 4, *Social support for health*, were significant predictors of MyHR engagement. A 1-unit increase in each scale was associated with 1.63-fold and 1.74-fold increase in the odds of engagement (Table 2 and Multimedia Appendix 4).

The use of the internet to search for health-related information continues to be an important predictor of MyHR engagement, in addition to awareness. Other strong predictors of engagement included age and the number of chronic conditions. Younger people were significantly more likely to engage with the MyHR, with every 10-year grouping increase being associated with a 20% reduction in the odds of MyHR engagement (OR 0.98, 95% CI 0.97-0.99). Having 2 or more chronic diseases or conditions was associated with a 1.88-fold increase in the odds of engagement compared with having no chronic conditions (OR 1.88, 95% CI 1.16-3.07; Multimedia Appendix 4).

### *Use or Intention to Use*

Digital health literacy was again a strong predictor of MyHR use or intention to use (Table 2 and Multimedia Appendix 5). A 1-unit increase in each of the eHLQ scales was associated with a 2.25- to 4.44-fold increase in the odds of using or intending to use MyHR, with scale 3, *Ability to engage with digital services*, being notably strong, followed by scale 5, *Motivated to engage with digital services*, and scale 1, *Using technology to process health information*.

Health literacy was also a significant predictor of MyHR use. A 1-unit increase in HLQ scale 1, *Feeling understood and supported by health care providers*, scale 3, *Actively managing my health*, and scale 4, *Social support for health*, was associated with a 1.89-, 2.28-, and 2.1-fold increase in the odds of using or intending to use MyHR.

Being female (OR 1.78, 95% CI 1.05-3.00), having a university education (OR 2.48, 95% CI 1.23-5.02), and using the internet to search for health-related information in the past year (OR 2.96, 95% CI 1.64-5.37) were more likely to use or intend to use MyHR compared with being male, not completing secondary school, or not using the internet to search for health-related information (Multimedia Appendix 5).

## Discussion

### Principal Findings

This study used a population-based survey to explore the predictors of awareness, engagement, and use of a national web-based PHR, MyHR, using multidimensional measures of health literacy. Digital health literacy was strongly and consistently associated with MyHR awareness, engagement, and use. Most notably, people who reported that they were using technology for health, had the ability and motivation to engage with digital services, and found that digital services met their individual needs were 3 to 4 times more likely to use or intend to use MyHR compared with their counterparts. Other clear associations included dimensions of health literacy relating to positive relationships with health care providers and social support and using the internet to search for health information.

### Comparison With Prior Work

The limited number of studies on the predictors of PHR uptake and the use of digital health literacy as a general concept in previous studies make it difficult to compare the findings of this study with those of prior studies. Nevertheless, this study found that only one-third of the sample (320/998, 32.1%) was engaged with MyHR, and only a very few participants (63/998, 6.3%) were current users. This is in line with the data revealed at an Australian Senate estimate hearing in December 2019 that only 4% of Australians logged into MyHR more than once [60].

In a qualitative study of 66 Australian women who were regular users of web-based health information, Lupton [61] found that factors such as lack of interest, security, and privacy concerns or not seeing any benefits of using MyHR were potential barriers to using MyHR. This study also identified eHLQ scale 4, *Feel safe and in control*, and scale 5, *Motivated to engage with digital services*, as strongly associated with MyHR awareness, engagement, and use or intention to use. Another notable finding is the eHLQ scale 2, *Understanding of health language and concepts*, which was not associated with awareness but was the strongest predictor of engagement and a moderately strong predictor of use. This finding echoes the October 2021 statistics from MyHR, which showed that the documents most viewed by Australians were pathology reports, with >1.6 million views, a huge jump of 613% compared with the views 12 months earlier [62]. The immunization report, introduced in early 2021, was also one of the top 10 items people looked at in October 2021, with 2.7 million views, an increase of 68% compared with views a month earlier [62]. This could indicate that a better understanding of one's health is one of the purposes for engaging with MyHR.

### Digital Health Literacy

An important finding of this study is the potential role of digital health literacy in the adoption and use of web-based PHR. Instead of simply reporting that higher or lower digital health literacy is linked to the uptake of PHR, as in the study by Noblin et al [33], this study examined the 7 dimensions of digital health literacy and the relative strengths of the association of digital health literacy with awareness and engagement of PHR. By using a multidimensional instrument, this study provided more nuanced insights into people's MyHR awareness, engagement, and use or intention to use. The dimensions of using technology for health (eHLQ scale 1, *Using technology to process health information*) and ability (eHLQ scale 3, *Ability to actively engage with digital services*) as well as motivation to use technology (eHLQ scale 5, *Motivated to engage with digital services*) are all strongly associated with using MyHR, with a 4.14- to 4.44-increase in the odds of using or intention to use web-based PHR. This is further confirmed, because using the internet to search for health information was another strong predictor, especially in use or intention to use.

When interpreting the results of this study, it is useful to consider the 7 dimensions of digital health literacy measured and ask the question, "How does a person develop in this dimension?" It is clear that a person's digital health literacy will develop along with their experiences of using digital health technologies; for better or worse, positive experiences will increase their trust, confidence, and perceptions of value, whereas negative experiences will do the opposite. Therefore, digital health literacy is clearly not just a predictor of the use of technologies but also, if not more so, a consequence, especially among people with few prior experiences of using digital health technologies.

However, the direction of causality was not assumed in this study. Although initiatives to develop digital skills may be an effective way to increase the adoption of MyHR (and is a commonly used strategy to improve digital health literacy [63]), an alternative is to use the assessment of health literacy as a means to understand and shape the experiences of people as they engage with digital health technologies and the processes that build (or undermine) trust, confidence, and perceived benefits. This would lead to a process of designing both the features and rolling out of digital technologies in such a way as to maximize virtuous cycles and minimize confidence eroding cycles, guided by an understanding of the digital health literacy of users. This overall process can be described as "health literacy development."

### Implications and Recommendations

Although digital skill training may have the potential to increase the likelihood of using web-based PHR and is likely to be practical and easy to implement, it should also be noted that other potential predictors identified in this study may also be important to consider when developing community-based interventions and leveraging the important role that health professionals may have in influencing people's knowledge and confidence in engaging with digital records. With eHLQ scale 4, *Feel safe and in control*, being a strong predictor, addressing people's privacy concerns is an action that may be a key building block to engagement. The eHLQ scale 2, *Understanding of health language and concepts*, while was not associated with being aware of MyHR, was moderately associated with engagement and use, suggesting that initiatives to promote the benefits of using the system to better understand and manage people's own health should be considered. In fact, it is likely that when people see the benefits of using web-based PHR, they may become more motivated to use digital technologies, leading to cyclical growth in both confidence and intent to use MyHR along with digital health literacy.

Besides, the notion of including other nondigital actions is further supported by the finding that predictors of engagement and use included the health literacy dimensions represented by the HLQ scale 1, *Feeling understood and supported by health care providers*, scale 3, *Actively managing my health*, and scale 4, *Social support for health*. This indicates that public health education should not only target individual users of MyHR but also provide supportive health care and social networks to encourage the use of web-based PHR. Hence, an approach that considers the social factors surrounding the use of web-based PHR is a critical aspect of health literacy development.

In a review of the information quality and usability of MyHR in 2018 using a health literacy framework, Walsh et al [64] found that only 16% of such resources could be rated as easy to read, 88% were text based, images to assist learning were limited, and color and large buttons to facilitate engagement and navigation were missing. They concluded that people at risk of lower health literacy did not have equitable access to the system, potentially increasing health disparities between users and nonusers. Although the MyHR website is regularly updated, the MyHR website as of March 2022 features a small button to listen to the website, and translation to other languages is available only for some pages, not the full website. As this study found that digital services that suit individual needs were associated with a 3.5-fold increase in the odds of using or intending to use MyHR, actions such as improving the readability, usability, and accessibility of MyHR to ensure that the system is responsive to all users' needs may maximize equitable access to web-based PHR. An easy-to-use digital system may boost people's confidence and motivation to use digital health technologies and, in turn, reinforce their engagement with the system and perhaps develop the digital health literacy of individuals.

Furthermore, this study found that older people and those with less education were less likely to engage in or use MyHR. These findings indicate that there are population groups that may become disadvantaged as countries move to web-based PHR, leading to a potential widening of health disparities. Special initiatives with targeted and tailored interventions are needed to ensure that no group remains left behind. A co-design approach to developing these initiatives with those with lived experience in diverse communities is also recommended, because this approach has been considered the best practice "to reduce inequality and empower vulnerable communities" [65].

## Future Directions

This study was the needs assessment phase of a larger study that used the Ophelia process to identify solutions to maximize the use of digital technologies, including MyHR. The Ophelia process takes a co-design approach to create and implement solutions to understand and improve access, equity, and outcomes by addressing health literacy needs [38]. It has been shown to be effective in various studies [66-69] and is easy to apply in diverse settings from European hospital settings to Egyptian fishing villages [70-73]. In addition to identifying the predictors of PHR use, the data of this study were also used to identify digital health literacy profiles among community members. These profiles reflect the digital health literacy challenges, preferences, and strengths of respondents regarding the adoption and use of MyHR. The next phase of this study involved integrating quantitative data profiles with demographic and interview data to create vignettes (or stories) of typical community members about their use of digital health systems. The vignettes were then presented at idea generation workshops attended by community members and frontline health professionals. This process allows for co-design opportunities using local wisdom to generate clear and actionable recommendations to increase the uptake and use of MyHR and achieve widespread participation in the process of health literacy development.

The finding that only 6.3% of the survey sample were MyHR users indicates that attempts to engage the Australian population with MyHR might not have been successful at that time. Since the opt-out period when around 10% (approximately 2.5 million) of Australians opted out of the system [74], approximately 90,000 people who previously opted out or canceled their records had registered for a record by June 2021 [75]. Further population-based surveys should be undertaken to measure changes in and predictors of awareness, engagement, and use of MyHR, especially as the use of digital health services has become more common during the COVID-19 pandemic. The findings will inform further actions to improve the uptake rate of active MyHR users.

## Limitations

A limitation of this study is that it was conducted in a regional area of Victoria, which may not be representative of the Australian population. However, the aim of this study was to identify digital health literacy needs, and rural residents were found to be among the population groups with greater digital health literacy needs than urban residents [19,21]. It is also noted that there were some demographic groups that were not well represented in the sample when compared with that of residents of the Ballarat Goldfields region; this included a lower representation of younger people and those with lower education. The impact of this discrepancy is that the findings may not be generalizable to the general adult population in the region. However, it is important to note that this study sought to understand people's experience with digital health services. Given that people in middle to late adulthood have the highest burden of chronic health conditions and tend to have multimorbidity, this population group may have more to gain from the engagement and use of web-based PHR.

Groups experiencing vulnerabilities who are most at risk of digital inequities were also not well represented. Only 1.8% (18/998) of the respondents identified as Aboriginal or Torres Strait Islanders and only 0.8% (8/998) spoke a language other than English at home. Although the proportion of Aboriginal and Torres Strait Islanders was similar to the population of the Ballarat Goldfields region (Multimedia Appendix 2), the proportion of people who spoke a language other than English at home was lower than that in the 2016 Australian Bureau of Statistics Census (39,743/576,802, 6.9%) [76]. Future studies should focus on these 2 groups to better understand their needs and barriers to digital health.

Another limitation is that the sample included only adults who had a landline telephone or mobile phone number registered to a postcode in the Ballarat Goldfields region. Consequently, individuals who did not have access to a phone, who lived in the region but had a phone registered to a different postcode, or who had an unlisted number were not sampled. Such individuals may have had different experiences with health technologies and services than the study sample, and their experiences were not captured in the data. Furthermore, CATI as an administration mode can lead to some cognitive burden and recall bias, which may in turn affect the ways in which people respond [77]. However, the use of a telephone survey is a cost-effective way to undertake a large population survey in a short period.

### Conclusions

This study provides insights into the predictors of the use of a national web-based PHR, MyHR, in Australia, and advances the understanding of the mechanism behind the use of web-based PHRs. These findings suggest that actions to improve the uptake and use of web-based PHRs need to look beyond improving individual digital skills. Of equal importance are initiatives to provide access to digital technologies, develop responsive digital services, provide a better understanding of the benefits of using web-based PHRs, and establish health care and social support networks. Therefore, a holistic approach is essential for enhancing the rate of web-based PHR engagement and use. This study also identified subgroups that are likely to be nonusers of web-based PHRs; targeted solutions need to be put in place to ensure that a web-based PHR can realize its full potential to help reduce health inequities.

---

## Acknowledgments

The authors would like to acknowledge the contribution of members of the community residing in the Western Victoria Primary Health Network (WVPHN) region who participated in this research. The authors also thank the staff of WVPHN for sharing their perspectives and experiences throughout the project. We acknowledge and thank the staff and management of Strahan Research Pty Ltd for the timeliness and quality of completion of the CATI process. This project was funded by WVPHN through a grant from the Australian Digital Health Agency. RHO received funds from the National Health and Medical Research Council of Australia Principal Research Fellowship APP1155125.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
eHealth Literacy Questionnaire and Health Literacy Questionnaire scales and definitions.
[DOCX File , 22 KB - jmir_v24i9e35772_app1.docx ]

Multimedia Appendix 2
Comparison of study participants and the population of the Ballarat Goldfields region of the Western Victoria Primary Health Network.
[DOCX File , 15 KB - jmir_v24i9e35772_app2.docx ]

Multimedia Appendix 3
Logistic regression predicting likelihood of My Health Record awareness.
[DOCX File , 90 KB - jmir_v24i9e35772_app3.docx ]

Multimedia Appendix 4
Logistic regression predicting likelihood of My Health Record engagement.
[DOCX File , 88 KB - jmir_v24i9e35772_app4.docx ]

Multimedia Appendix 5
Logistic regression predicting likelihood of My Health Record use or intention to use.
[DOCX File , 84 KB - jmir_v24i9e35772_app5.docx ]

## References

1. Alsahafi AY, Gay BV. An overview of electronic personal health records. Health Policy Technol 2018 Dec;7(4):427-432. [doi: 10.1016/j.hlpt.2018.10.004]
2. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. BMJ 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]

3. Hilton JF, Barkoff L, Chang O, Halperin L, Ratanawongsa N, Sarkar U, et al. A cross-sectional study of barriers to personal health record use among patients attending a safety-net clinic. PLoS One 2012 Feb 20;7(2):e31888 [FREE Full text] [doi: 10.1371/journal.pone.0031888] [Medline: 22363761]
4. Dutta B, Peng M, Sun S. Modeling the adoption of personal health record (PHR) among individual: the effect of health-care technology self-efficacy and gender concern. Libyan J Med 2018 Dec 24;13(1):1500349 [FREE Full text] [doi: 10.1080/19932820.2018.1500349] [Medline: 30037314]
5. Global Diffusion of EHealth: Making Universal Health Coverage Achievable Report of the Third Global Survey on EHealth. Geneva, Switzerland: World Health Organization; 2017.
6. Global Strategy on Digital Health 2020-2025. Geneva, Switzerland: World Health Organization; 2021.
7. What can My Health Record data be used for? Australian Digital Health Agency. URL: https://www.myhealthrecord.gov.au/for-you-your-family/howtos/choose-how-your-data-is-used-for-research [accessed 2022-06-03]
8. Lehnbom EC, Brien JE, McLachlan AJ. Knowledge and attitudes regarding the personally controlled electronic health record: an Australian national survey. Intern Med J 2014 Apr;44(4):406-409. [doi: 10.1111/imj.12384] [Medline: 24754689]
9. My Health Record a potential life saver for Australians from culturally and linguistically diverse backgrounds. Australian Digital Health Agency. URL: https://www.myhealthrecord.gov.au/news-and-media/media-releases/life-saver-for-culturally-linguistically-diverse-backgrounds [accessed 2020-06-06]
10. Walsh L, Hemsley B, Allan M, Adams N, Balandin S, Georgiou A, et al. The e-health literacy demands of Australia's my health record: a heuristic evaluation of usability. Perspect Health Inf Manag 2017;14(Fall):1f [FREE Full text] [Medline: 29118683]
11. My Health Record opt out date announced. Australian Government Department of Health. URL: https://www.myhealthrecord.gov.au/news-and-media/media-releases/my-health-record-opt-out-date-announced [accessed 2020-06-03]
12. Gaylin DS, Moiduddin A, Mohamoud S, Lundeen K, Kelly JA. Public attitudes about health information technology, and its relationship to health care quality, costs, and privacy. Health Serv Res 2011 Jun;46(3):920-938 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01233.x] [Medline: 21275986]
13. The devil’s in the detail: final report of the independent evaluation of the Summary Care Record and Healthspace programmes. NHS Networks. URL: https://www.networks.nhs.uk/news/the-devil2019s-in-the-detail-final-report-of-the-independent-evaluation-of-the-summary-care-record-and-healthspace-programmes [accessed 2021-07-16]
14. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
15. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
16. Monkman H, Kushniruk A. eHealth literacy issues, constructs, models, and methods for health information technology design and evaluation. Knowl Manag E Learn 2015;7(4):541-549. [doi: 10.34105/j.kmel.2015.07.036]
17. Gibbons MC, Casale CR. Reducing disparities in health care quality: the role of health IT in underresourced settings. Med Care Res Rev 2010 Oct;67(5 Suppl):155S-1562. [doi: 10.1177/1077558710376202] [Medline: 20829238]
18. Healthcare disparities and the role of personal health records. The American Health Information Management Association. URL: https://library.ahima.org/doc?oid=91677#.YtL8GHZBzIU [accessed 2021-04-10]
19. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
20. Litt E. Measuring users’ internet skills: a review of past assessments and a look toward the future. New Media Soc 2013 May 24;15(4):612-630. [doi: 10.1177/1461444813475424]
21. Falling through the net: Defining the digital divide. Washington, D. C. National Telecommunications and Information Administration. URL: https://www.ntia.doc.gov/legacy/ntiahome/fttn99/FTTN.pdf [accessed 2021-04-10]
22. Wei K, Teo H, Chan HC, Tan BC. Conceptualizing and testing a social cognitive model of the digital divide. Inform Syst Res 2011 Mar;22(1):170-187. [doi: 10.1287/isre.1090.0273]
23. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
24. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep 29;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
25. Lustria ML, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243 [FREE Full text] [doi: 10.1177/1460458211414843] [Medline: 21937464]
26. van Deursen AJ, van Dijk JA, Peters O. Rethinking internet skills: the contribution of gender, age, education, internet experience, and hours online to medium- and content-related internet skills. Poetics 2011 Apr;39(2):125-144. [doi: 10.1016/j.poetic.2011.02.001]

27. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
28. Gerber DE, Laccetti AL, Chen B, Yan J, Cai J, Gates S, et al. Predictors and intensity of online access to electronic medical records among patients with cancer. J Oncol Pract 2014 Sep;10(5):e307-e312 [FREE Full text] [doi: 10.1200/JOP.2013.001347] [Medline: 25006222]
29. Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WS, Patel V, Moser RP, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017 Jun 02;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
30. Kim E, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009 Oct 27;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
31. Kim S, Abner E. Predictors affecting personal health information management skills. Inform Health Soc Care 2016;41(3):211-229 [FREE Full text] [doi: 10.3109/17538157.2015.1008490] [Medline: 26268728]
32. Strekalova YA. Electronic health record use among cancer patients: insights from the Health Information National Trends Survey. Health Informatics J 2019 Mar 23;25(1):83-90 [FREE Full text] [doi: 10.1177/1460458217704246] [Medline: 28434278]
33. Noblin AM, Wan TT, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13 [FREE Full text] [Medline: 23209454]
34. McPhee C, Batterham R, Hawkins M, Gearon E, Hanna L, Osborne R. eHealth literacy study - Ballarat: Technical Report. Deakin University. 2019. URL: https://www.ausdoc.com.au/sites/default/files/e-health-literacy-study_2019.pdf [accessed 2021-03-28]
35. National, state and territory population. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release [accessed 2022-03-03]
36. Final Report My Health Record system. Canberra: Commonwealth of Australia; 2018.
37. Legislation and governance. Australian Digital Health Agency. URL: https://www.myhealthrecord.gov.au/about/legislation-and-governance [accessed 2021-03-12]
38. Batterham RW, Buchbinder R, Beauchamp A, Dodson S, Elsworth GR, Osborne RH. The OPtimising HEalth LIterAcy (Ophelia) process: study protocol for using health literacy profiling and community engagement to create and implement health reform. BMC Public Health 2014 Jul 07;14:694 [FREE Full text] [doi: 10.1186/1471-2458-14-694] [Medline: 25002024]
39. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016. Australian Bureau of Statistics. URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012016?OpenDocument [accessed 2018-09-28]
40. 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 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
41. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
42. Buchbinder R, Batterham R, Elsworth G, Dionne CE, Irvin E, Osborne RH. A validity-driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. Arthritis Res Ther 2011;13(5):R152 [FREE Full text] [doi: 10.1186/ar3468] [Medline: 21933393]
43. Elsworth GR, Beauchamp A, Osborne RH. Measuring health literacy in community agencies: a Bayesian study of the factor structure and measurement invariance of the health literacy questionnaire (HLQ). BMC Health Serv Res 2016 Sep 22;16(1):508 [FREE Full text] [doi: 10.1186/s12913-016-1754-2] [Medline: 27659559]
44. Debussche X, Lenclume V, Balcou-Debussche M, Alakian D, Sokolowsky C, Ballet D, et al. Characterisation of health literacy strengths and weaknesses among people at metabolic and cardiovascular risk: validity testing of the Health Literacy Questionnaire. SAGE Open Med 2018;6:2050312118801250 [FREE Full text] [doi: 10.1177/2050312118801250] [Medline: 30319778]
45. Kolarcik P, Cepova E, Madarasova Geckova A, Elsworth GR, Batterham RW, Osborne RH. Structural properties and psychometric improvements of the Health Literacy Questionnaire in a Slovak population. Int J Public Health 2017 Jun 3;62(5):591-604. [doi: 10.1007/s00038-017-0945-x] [Medline: 28258403]
46. Maindal HT, Kayser L, Norgaard O, Bo A, Elsworth GR, Osborne RH. Cultural adaptation and validation of the Health Literacy Questionnaire (HLQ): robust nine-dimension Danish language confirmatory factor model. Springerplus 2016;5(1):1232 [FREE Full text] [doi: 10.1186/s40064-016-2887-9] [Medline: 27536516]
47. Morris RL, Soh S, Hill KD, Buchbinder R, Lowthian JA, Redfern J, et al. Measurement properties of the Health Literacy Questionnaire (HLQ) among older adults who present to the emergency department after a fall: a Rasch analysis. BMC Health Serv Res 2017 Aug 29;17(1):605 [FREE Full text] [doi: 10.1186/s12913-017-2520-9] [Medline: 28851344]

48. Nolte S, Osborne RH, Dwinger S, Elsworth GR, Conrad ML, Rose M, et al. German translation, cultural adaptation, and validation of the Health Literacy Questionnaire (HLQ). PLoS One 2017;12(2):e0172340 [FREE Full text] [doi: 10.1371/journal.pone.0172340] [Medline: 28234987]
49. Hawkins M, Elsworth GR, Osborne RH. Application of validity theory and methodology to patient-reported outcome measures (PROMs): building an argument for validity. Qual Life Res 2018 Jul;27(7):1695-1710 [FREE Full text] [doi: 10.1007/s11136-018-1815-6] [Medline: 29464456]
50. 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 Oct 14;23(10):e30243 [FREE Full text] [doi: 10.2196/30243] [Medline: 34647897]
51. Cheng C, Elsworth G, Osborne R. Validity evidence of the eHealth literacy questionnaire (eHLQ) part 2: mixed methods approach to evaluate test content, response process, and internal structure in the Australian community health setting. J Med Internet Res 2022 Mar 08;24(3):e32777 [FREE Full text] [doi: 10.2196/32777] [Medline: 35258475]
52. Hawkins M, Gill SD, Batterham R, Elsworth GR, Osborne RH. The Health Literacy Questionnaire (HLQ) at the patient-clinician interface: a qualitative study of what patients and clinicians mean by their HLQ scores. BMC Health Serv Res 2017 Apr 27;17(1):309 [FREE Full text] [doi: 10.1186/s12913-017-2254-8] [Medline: 28449680]
53. Goodwin B, March S, Zajdlewicz L, Osborne R, Dunn J, Chambers S. Health literacy and the health status of men with prostate cancer. Psychooncology 2018 Oct;27(10):2374-2381 [FREE Full text] [doi: 10.1002/pon.4834] [Medline: 29959807]
54. Budhathoki SS, Hawkins M, Elsworth G, Fahey MT, Thapa J, Karki S, et al. Use of the English health literacy questionnaire (HLQ) with health science university students in Nepal: a validity testing study. Int J Environ Res Public Health 2022 Mar 09;19(6):3241 [FREE Full text] [doi: 10.3390/ijerph19063241] [Medline: 35328928]
55. Wahl AK, Hermansen Å, Osborne RH, Larsen MH. A validation study of the Norwegian version of the Health Literacy Questionnaire: a robust nine-dimension factor model. Scand J Public Health 2021 Jun;49(4):471-478 [FREE Full text] [doi: 10.1177/1403494820926428] [Medline: 32508258]
56. Boateng MA, Agyei-Baffour P, Angel S, Enemark U. Translation, cultural adaptation and psychometric properties of the Ghanaian language (Akan; Asante Twi) version of the Health Literacy Questionnaire. BMC Health Serv Res 2020 Nov 23;20(1):1064 [FREE Full text] [doi: 10.1186/s12913-020-05932-w] [Medline: 33228648]
57. Rademakers J, Waverijn G, Rijken M, Osborne R, Heijmans M. Towards a comprehensive, person-centred assessment of health literacy: translation, cultural adaptation and psychometric test of the Dutch Health Literacy Questionnaire. BMC Public Health 2020 Dec 02;20(1):1850 [FREE Full text] [doi: 10.1186/s12889-020-09963-0] [Medline: 33267834]
58. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
59. Primary health networks. Australian Government Department of Health and Aged Care. URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/PHN-Demographic_Data [accessed 2019-01-08]
60. McCauley D. Australians shun My Health Record with only 9 per cent ever logging in. The Sydney Morning Herald. URL: https://www.smh.com.au/politics/federal/australians-shun-my-health-record-with-only-9-per-cent-ever-logging-in-20191220-p53lz0.html [accessed 2021-04-10]
61. Lupton D. 'I'd like to think you could trust the government, but I don't really think we can': Australian women's attitudes to and experiences of My Health Record. Digit Health 2019 Apr 29;5:2055207619847017 [FREE Full text] [doi: 10.1177/2055207619847017] [Medline: 31069106]
62. Statistics and Insights OCTOBER 2021. My Health Record. URL: https://www.digitalhealth.gov.au/sites/default/files/2021-11/mhr-stats-october.pdf [accessed 2021-12-11]
63. Pourrazavi S, Kouzekanani K, Bazargan-Hejazi S, Shaghaghi A, Hashemiparast M, Fathifar Z, et al. Theory-based E-health literacy interventions in older adults: a systematic review. Arch Public Health 2020 Aug 10;78(1):72 [FREE Full text] [doi: 10.1186/s13690-020-00455-6] [Medline: 32793345]
64. Walsh L, Hemsley B, Allan M, Dahm MR, Balandin S, Georgiou A, et al. Assessing the information quality and usability of within a health literacy framework: what's changed since 2016? Health Inf Manag 2021;50(1-2):13-25 [FREE Full text] [doi: 10.1177/1833358319864734] [Medline: 31370712]
65. Goodyear-Smith F, Jackson C, Greenhalgh T. Co-design and implementation research: challenges and solutions for ethics committees. BMC Med Ethics 2015 Nov 16;16:78 [FREE Full text] [doi: 10.1186/s12910-015-0072-2] [Medline: 26573410]
66. O'Hara J, McPhee C, Dodson S, Cooper A, Wildey C, Hawkins M, et al. Barriers to breast cancer screening among diverse cultural groups in Melbourne, Australia. Int J Environ Res Public Health 2018 Aug 07;15(8):1677 [FREE Full text] [doi: 10.3390/ijerph15081677] [Medline: 30087259]
67. Beauchamp A, Mohebbi M, Cooper A, Pridmore V, Livingston P, Scanlon M, et al. The impact of translated reminder letters and phone calls on mammography screening booking rates: two randomised controlled trials. PLoS One 2020;15(1):e0226610 [FREE Full text] [doi: 10.1371/journal.pone.0226610] [Medline: 31923178]
68. Beauchamp A, Batterham RW, Dodson S, Astbury B, Elsworth GR, McPhee C, et al. Systematic development and implementation of interventions to OPtimise Health Literacy and Access (Ophelia). BMC Public Health 2017 Mar 03;17(1):230 [FREE Full text] [doi: 10.1186/s12889-017-4147-5] [Medline: 28253883]

69. Goeman D, Conway S, Norman R, Morley J, Weerasuriya R, Osborne RH, et al. Optimising health literacy and access of service provision to community dwelling older people with diabetes receiving home nursing support. J Diabetes Res 2016;2016:2483263 [FREE Full text] [doi: 10.1155/2016/2483263] [Medline: 27668261]
70. Aaby A, Simonsen CB, Ryom K, Maindal HT. Improving organizational health literacy responsiveness in cardiac rehabilitation using a co-design methodology: results from the heart skills study. Int J Environ Res Public Health 2020 Feb 05;17(3):1015 [FREE Full text] [doi: 10.3390/ijerph17031015] [Medline: 32033503]
71. Bakker M, Putrik P, Aaby A, Debussche X, Morrissey J, Råheim BC. Acting together – WHO National Health Literacy Demonstration Projects ( NHLDPs)  address health literacy needs in the European Region. Public Health Panorama 2019;5(2-3) [FREE Full text]
72. Cheng C, Elsworth GR, Osborne RH. Co-designing eHealth and Equity Solutions: application of the Ophelia (Optimizing Health Literacy and Access) Process. Front Public Health 2020;8:604401 [FREE Full text] [doi: 10.3389/fpubh.2020.604401] [Medline: 33330344]
73. Anwar WA, Mostafa NS, Hakim SA, Sos DG, Cheng C, Osborne RH. Health literacy co-design in a low resource setting: harnessing local wisdom to inform interventions across fishing villages in egypt to improve health and equity. Int J Environ Res Public Health 2021 Apr 24;18(9):4518 [FREE Full text] [doi: 10.3390/ijerph18094518] [Medline: 33923187]
74. My Health opt-outs jump to 2.5 million. Special Broadcasting Service Australia. URL: https://www.sbs.com.au/news/article/my-health-opt-outs-jump-to-2-5-million/ldfuy3act [accessed 2021-08-23]
75. Statistics and Insights. The Australian Digital Health Agency. URL: https://www.myhealthrecord.gov.au/sites/default/files/mhr_stats_marchnovember2019.pdf?v=1576471841 [accessed 2021-08-23]
76. Western Victoria Primary Health Network Needs Assessment 2017. Western Victoria Primary Health Network. URL: https://westvicphn.com.au/wp-content/uploads/2019/10/WVPHN_Needs_Assessment_2017_Core.pdf [accessed 2022-03-07]
77. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-291. [doi: 10.1093/pubmed/fdi031] [Medline: 15870099]

## Abbreviations

**CATI:** computer-assisted telephone interview
**eHLQ:** eHealth Literacy Questionnaire
**HLQ:** Health Literacy Questionnaire
**MyHR:** My Health Record
**Ophelia:** Optimizing Health Literacy and Access
**OR:** odds ratio
**PHR:** personal health record
**WVPHN:** Western Victoria Primary Health Network

*Edited by T Leung; submitted 17.12.21; peer-reviewed by T Greenhalgh, H Fang, A Barker; comments to author 15.02.22; revised version received 16.05.22; accepted 20.06.22; published 16.09.22.*

*Please cite as:*
*Cheng C, Gearon E, Hawkins M, McPhee C, Hanna L, Batterham R, Osborne RH*
*Digital Health Literacy as a Predictor of Awareness, Engagement, and Use of a National Web-Based Personal Health Record: Population-Based Survey Study*
*J Med Internet Res 2022;24(9):e35772*
*URL: https://www.jmir.org/2022/9/e35772*
*doi:10.2196/35772*
*PMID:36112404*

Original Paper

# Business Process Model and Notation and openEHR Task Planning for Clinical Pathway Standards in Infections: Critical Analysis

Natalia Iglesias[1], PhD; Jose M Juarez[1], PhD; Manuel Campos[1,2], PhD

[1]Instituto de Investigación de Tecnologías de la Información y las Comunicaciones Orientadas, University of Murcia, Murcia, Spain

[2]Instituto Murciano de Investigación Biosanitaria - Arrixaca, Murcia, Spain

**Corresponding Author:**
Natalia Iglesias, PhD
Instituto de Investigación de Tecnologías de la Información y las Comunicaciones Orientadas
University of Murcia
Faculty of Computer Science, Campus Espinardo
University of Murcia
Murcia, 30100
Spain
Phone: 34 868887864
Email: natalia.iglesias@um.es

## Abstract

**Background:** Clinical pathways (CPs) are usually expressed by means of workflow formalisms, providing health care personnel with an easy-to-understand, high-level conceptual model of medical steps in specific patient conditions, thereby improving overall health care process quality in clinical practice. From a standardized perspective, the *business process model and notation* (BPMN), a widely spread general-purpose process formalism, has been used for conceptual modeling in clinical domains, mainly because of its easy-to-use graphical notation, facilitating the common understanding and communication of the parties involved in health care. However, BPMN is not particularly oriented toward the peculiarities of complex clinical processes such as infection diagnosis and treatment, in which time plays a critical role, which is why much of the BPMN clinical-oriented research has revolved around how to extend the standard to address these special needs. The shift from an agnostic, general-purpose BPMN notation to a natively clinical-oriented notation such as *openEHR Task Planning* (TP) could constitute a major step toward clinical process improvement, enhancing the representation of CPs for infection treatment and other complex scenarios.

**Objective:** Our work aimed to analyze the suitability of a clinical-oriented formalism (TP) to successfully represent typical process patterns in infection treatment, identifying domain-specific improvements to the standard that could help enhance its modeling capabilities, thereby promoting the widespread adoption of CPs to improve medical practice and overall health care quality.

**Methods:** Our methodology consisted of 4 major steps: identification of *key features* of infection CPs through literature review, clinical guideline analysis, and BPMN extensions; analysis of the presence of *key features* in TP; modeling of relevant process patterns of catheter-related bloodstream infection as a case study; and analysis and proposal of extensions in view of the results.

**Results:** We were able to easily represent the same logic applied in the extended BPMN-based process models in our case study using *out-of-the-box* standard TP primitives. However, we identified possible improvements to the current version of TP to allow for simpler conceptual models of infection CPs and possibly of other complex clinical scenarios.

**Conclusions:** Our study showed that the clinical-oriented TP specification is able to successfully represent the most complex catheter-related bloodstream infection process patterns depicted in our case study and identified possible extensions that can help increase its adequacy for modeling infection CPs and possibly other complex clinical conditions.

***(J Med Internet Res 2022;24(9):e29927)*** doi:10.2196/29927

**KEYWORDS**

openEHR task planning; business process model and notation; BPMN; clinical pathways; catheter-related bloodstream infection; CR-BSI; clinical guidelines

## Introduction

### Background

The interoperability of clinical information is a major issue that hinders data exchange and knowledge reuse between clinical institutions [1]. An important type of clinical information is the representation of institution-specific care protocols, known as clinical pathways (CPs), defined as "task orientated care plans which detail essential steps in the care of patients with a specific clinical problem and describe the patient's expected clinical course" [2]. This definition has been reviewed in the studies by De Bleser et al [3] and Vanhaecht et al [4] and further refined in the study by Kinsman et al [5] through five criteria: "(1) a structured multidisciplinary plan of care; (2) used to channel the translation of guidelines or evidence into local structures; (3) detailed the steps in a course of treatment or care in a plan, pathway, algorithm, guideline, protocol or other inventory of actions; (4) had timeframes or criteria-based progression (that is, steps were taken if designated criteria were met); and (5) aimed to standardise care for a specific clinical problem, procedure or episode of healthcare in a specific population." CPs often rely on medical evidence expressed in clinical guidelines (CGs), which standardize abstract best practices for the diagnosis and treatment of specific medical conditions based on both medical evidence and expert consensus to ultimately improve the quality and uniformity of care, describing the strict temporal order in which clinical work needs to be carried out. For this purpose, many clinical task-oriented tools have been used over the years, such as Asbru, GLIF, GLARE, PROforma, EON, or GUIDE [6], but none of them seem to have reached the popularity of general-purpose process management tools such as the *business process model and notation* (BPMN) from the Object Management Group (OMG) [7], probably because of their complexity and narrow use limited to high-technology institutions with enough financial and technical means. In the last years, BPMN, which is widely used in other industry domains, has also emerged in clinical domains as a process management standard. Many studies describe the use of BPMN to represent clinical processes to improve efficiency or serve as a basis for the development of clinical decision support systems, which require seamless integration between electronic health record (EHR) data, decisions, and the specific workflow of a medical institution [8-12]. Although BPMN has proven to be effective in representing clinical processes, the transition to process execution in real clinical institutions is still scarce [13]. Indeed, although BPMN can effectively help clinicians visually understand clinical processes and detect possible inefficiencies, the implementation of otherwise complex, stepwise clinical workflows in an ordered manner is not easy to accomplish as temporal interrelations between tasks are crucial. In the last years, new clinical-oriented process representation standards have emerged that address the specific needs of clinical workflows, such as openEHR Task Planning (TP) [14]. Our work is based on the hypothesis that the clinical-oriented business process management (BPM) tool TP provides a more adequate way to represent CPs, bridging the gap between abstract CG logic and clinical workflow and offering essential tools for the representation of complex health care processes such as infection treatment. In our research, we analyze how TP can represent complex infection CPs taking as a case study the catheter-related bloodstream infection (CR-BSI) CGs developed by the Johns Hopkins Hospital (JHH) Antimicrobial Stewardship Program, an international *gold standard* that includes recommendations aiming to standardize clinical practice around antibiotic prescription, thereby helping minimize antibiotic resistance [15].

The structure of this paper is as follows: in the *Introduction* section, we explain the background; in the *Methods* section, we illustrate the methodology used; in the *Results* section, we describe our experiments; and, in the *Discussion* section, we discuss the relevance and limitations of the results.

### Related Work

The best up-to-date evidence-based clinical knowledge is usually expressed in CGs, defined as "systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances" [16], often used in a paper-based mode. Many studies have successfully represented CG knowledge using different notations, such as the procedural medical-oriented *Arden Syntax*, rule-based systems such as *Drools*, guideline definition languages (*Graphic Language for Interactive Design [GLIDE]* and the *openEHR Guideline Definition Language*), or *Semantic Web* rule systems such as *SPARQL Inferencing Notation* (SPIN) or *Shapes Constraint Language* (SHACL) [17]. However, knowledge representation technologies do not suffice to represent a specific patient evolution over time, which has led to the development of medical-oriented task-based systems such as *PROforma* [18], *Asbru* [19], or *Prodigy* [20], which have been successfully used in many clinical scenarios but are effectively limited to a few medical institutions owing to the high costs and efforts associated with their implementation. This has favored the introduction in the last years of popular, easy-to-use, general-purpose BPM standards in the health care landscape as a working alternative to complex, medical task-oriented knowledge systems. However, modeling specific constraints in clinical processes remains a challenge because of the intrinsic domain complexity. Much research has been conducted over the years to eliminate or compensate for the temporal shortcomings of generic BPM notations, such as with *Petri nets* (a graphical workflow formalism for conceptual modeling of distributed systems [21,22]), *Process Mining for Healthcare* [23], or improved integration with data or decision support [9,10,24].

### Clinical Context

CR-BSI is a highly recurring infection and a major cause of morbidity and costs in hospitals. The leading cause of CR-BSI is gram-positive bacteria present in intravascular catheters, especially the coagulase-negative *Staphylococcus* species, which must be treated with a multidisciplinary approach based on catheter removal or catheter salvage combined with an antimicrobial *lock therapy*. Only in the United States, >150 million intravascular catheters are purchased by hospitals each year [25,26]. In addition, >250,000 intravascular CR-BSI cases happen each year, with an attributed mortality rate of 12% to 25% [27]. Of these, approximately 80,000 CR-BSI cases take place in intensive care units, where “more than 15 million central vascular catheter (CVC) days occur each year” [28]. This happens in a context of insufficient research on new antibiotics and limited supply of the existing ones, which narrows down the possible therapies for highly recurring bacterial infections, many of which are still susceptible to generic antibiotics with lower toxicity levels and lower risk of resistance development [29]. The treatment options for CR-BSI depend on the microorganism causing the infection, which determines the type of antibiotic and the time, frequency, and dosage of its intake. Furthermore, the type of catheter, the way it is handled, the duration of its placement, and the patient conditions codetermine the risk of a hospital-acquired CR-BSI, increasing the length of hospital stay and mortality figures especially in patients who are critically ill and whose catheter is not removed [30]. Therefore, specific CGs have been developed for health care staff to prevent the development of CR-BSI and issue recommendations regarding the most suitable course of treatment. The JHH CGs, based on current literature, Infectious Diseases Society of America national guidelines [31], and JHH medical evidence, cover the entire infection course from diagnosis through antibiotic treatment and assist clinicians in the selection of the optimal antibiotic therapy in an attempt to fight antibiotic resistance and still remain effective [15].

### BPMN Process Formalism

BPMN is a mature, general-purpose BPM graphical representation and ISO standard developed by OMG based on an unstructured graph-oriented language combined with features from other workflow languages that can be represented using *Petri nets* [32]. In BPMN, a process is a free sequence of activities or events ordered in a sequence flow and connected through split or merge gateways that redirect the flow into one or multiple paths. This standard has been widely used by business process managers in many different application domains owing to its simplicity. Despite not being specifically designed for clinical processes, BPMN has proven its value in the health care domain to represent CPs, allowing for an easy-to-understand representation of CP recommendations [12,13,33,34]. However, its use in clinical domains is still scarce owing to the complexity of clinical processes and organizations, the critical relevance of care protocols, the difficulties in handling temporal constraints, and the uncertainty surrounding patient evolution and treatment effectiveness [13]. Furthermore, the integration of BPMN process models with EHR data is usually modeled using Unified Modeling Language [35] to address the BPMN shortcomings in that regard [24]. In addition, BPMN provides a limited set of capabilities for modeling temporal constraints [34,36] as the standard lacks out-of-the-box temporal semantics. Therefore, many extensions of the BPMN metadata model have been proposed over the years to address the specific demands of clinical processes [11,34,36-44].

### openEHR TP

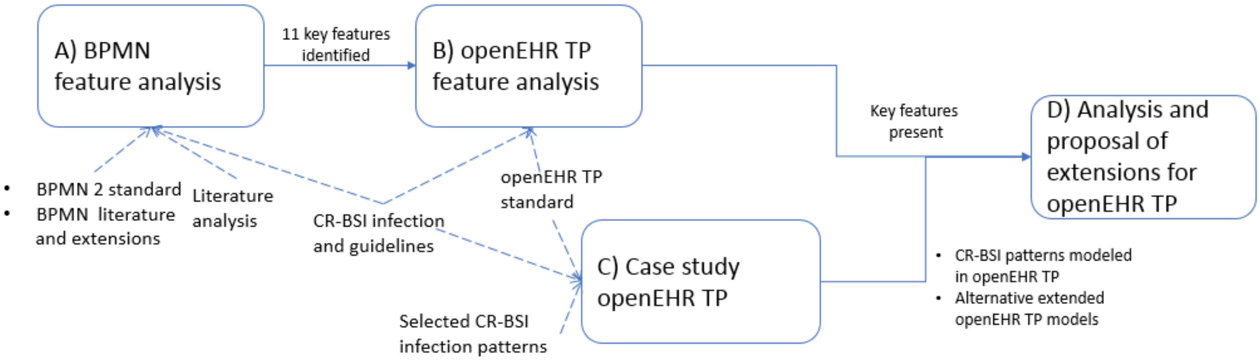
openEHR is an initiative of the openEHR Foundation working closely with the European Committee for Standardization (CEN), the International Organization for Standardization (ISO), Health Level 7 (HL7), OMG, and other organizations on EHR and clinical modeling standards that provides an archetype-based standard data model for EHR systems designed with interoperability in mind [45]. In 2016, the openEHR Foundation released a new clinical-oriented workflow management standard, the TP specification, that includes a Visual Modeling Language (TP-Visual Modeling Language [TP-VML]) backed up by formal semantics, which ultimately would allow for the automatic translation of graphical workflow models into executable models [14,46]. TP complements openEHR by allowing modeling orders and actions in the future, structured as work or task plans, as well as their eventual execution in distributed environments, improving features partially present in other workflow languages (BPMN, Yet Another Workflow Language [YAWL] [47], Case Management Model and Notation [CMMN] [48], or Decision Model and Notation [DMN] [49]). The TP engine executes both TP work plans and openEHR Decision Language (DL) rules [50]. Moreover, TP allows for the specification of complex times in medication orders (eg, “3 times a day before meals” [51]), and it is conceived as a clinical process navigator, empowering users to perform ad hoc modifications at the time of execution to reflect real-time changes and supporting auditing, reporting, and billing to analyze task performance, execution, and eventual deviations from the original plan. Although openEHR has a wide coverage as a research area [52-54], the TP specification has been used in fewer studies [55-57] and is still under development. The contributions of our work to support our hypothesis of the native suitability of clinical-oriented BPM formalisms for infection CPs are as follows: (1) identification of key features required for modeling infection CPs, as laid out in the literature, the JHH CGs, and the successive BPMN extensions; (2) theoretical analysis of how TP addresses the identified key features; (3) a case study to empirically show the native TP adequacy by modeling typical CR-BSI infection patterns using TP; and (4) analysis and proposal of potential extensions of the current version of TP for improved modeling of infection CPs and possibly of other complex CPs.

## Methods

### Overview

Our research aimed to explore and identify *key features* to represent infection CPs to then analyze how they are present in the TP standard. The analysis can lead to a potential list of benefits and possible improvements for the representation of infection CPs using TP. For this purpose, in this study we adopt the methodology illustrated in Figure 1.

**Figure 1.** Methodology followed. BPMN: business process model and notation; CR-BSI: catheter-related bloodstream infection; TP: Task Planning; TP-VML: Task Planning Visual Modeling Language.



### Step A

We identified *key features* required for modeling infection CPs by reviewing scientific literature, BPMN extensions in clinical settings, JHH CGs, and an *extended BPMN*-based model of JHH CR-BSI. The identified *key features* formed a basic set of requirements that served as a starting point to analyze the theoretical adequacy of TP for the representation of infection CPs.

### Step B

We conducted a theoretical analysis of the presence of the previously identified *key features* in the TP specification by performing an in-depth evaluation of the current version of TP to understand how these capabilities were addressed by the standard. When possible, a one-to-one comparison of both standards' capabilities was performed, which could help identify possible improvements. In addition, when new, potentially interesting TP capabilities were detected, we discussed their relevance.

### Step C

We conducted an empirical test of the suitability of the current version of TP for conceptual modeling of infection CPs. To that end, we modeled typical infection treatment patterns from the JHH CR-BSI CGs using the same logic used in the extended BPMN version [44]. When opportunities for improvements were detected, we proposed alternative TP models.

### Step D

Taking the output of steps B and C, we proceeded with the analysis of benefits and potential enhancements to the current version of the TP standard and made a proposal of possible extensions to improve the representation of infection CPs. For this purpose, we followed the steps of the methodology for BPMN extensions laid out in the study by Braun and Schlieter [58], which we consider adequate for TP.

## Results

### Step A: Identification of Key Features for Infection CPs

BPMN has been extended over the years with domain-specific concepts to meet the main requirements of process representations in clinical domains. Therefore, an analysis of the BPMN missing capabilities through its extensions in clinical settings is essential to identify *key features* that any CP representation should fulfill. Furthermore, any language for modeling CPs should contain (1) concepts for medical business process modeling (*patient state*, *treatment step*, *decision*, or *process flow*) plus the ability to integrate information objects and responsibilities, (2) indefinite order relations as well as compulsory parallel relations between treatment steps and iterating treatment steps, (3) the evidence class of any recommendation and decision and a link to the source of the evidence, and (4) temporal dependencies and explicit time events [58].

The *key features* proposed in this study were derived from an analysis of existing literature selected based on quality and content regarding (1) BPM extensions in clinical domains affecting the representation of, among others, time, resources, and coordination of CPs by multidisciplinary teams working together in a care process [8,36-39,41,58-63]; (2) review studies of BPMN extensions both in clinical and nonclinical domains [13]; and (3) analysis of JHH CG features for infection management [15,43,44], such as parallel work, synchronization with EHR data, or uncertainty in the course of an infection treatment. Although broadly inclusive, our criteria were proposed as a set of minimum requirements to represent infection CPs. First, the *well-structuredness* of a process conceptual model helps increase its legibility and understanding by clinical staff, thereby enabling the detection and analysis of possible process improvements. In addition, it is a requirement to avoid deadlocks and inconsistencies in and between processes, thereby facilitating the future maintenance and evolution of the process representation. *Process modularity* allows for the decomposition of large, complex clinical processes with many involved parties in their constituent subprocesses, thereby allowing for increased readability, process reuse, and faster model building and verification of soundness across the organization. *Events* allow for conceptual modeling of facts that take place at the time of execution, the occurrence of which is possibly unknown at the time of design. Events deal with uncertainty and are always present in the treatment of infections, whose course depends on factors unknown at the time of design, such as the patients' eventual response to medication. *Parallel execution of tasks* optimizes execution time to promptly react to clinical happenings as clinical institutions consist of complex multilayered organizational units, possibly externalized and interacting with each other. *Task duration* allows for the expression of a constraint, a deadline, or simply information to all process actors. For example, a blood test can take a minimum of 1 day and a maximum of 2 days to complete. If purely informational, workflow execution should not wait for this task to complete but, if meant as a constraint, the model should include semantics to hold the execution of subsequent tasks. The time distance between nonconsecutive tasks can be specified using *relative time constraints* such as edge duration between the starting or ending instant of a predecessor task A and the starting or ending instant of a successor task B. These constraints are determined by best practices or the institution's average figures to define a minimum or maximum waiting time when dispatching work to other organizational units. Synchronization between subprocesses is ultimately reduced to the synchronization between their inner tasks. *Use of resources* is relevant in infection CPs as catheters are a main cause of infection in hospitalized patients, requiring both *catheter lock therapy (CLT)* and *systemic therapy (ST)* applied repeatedly and alternatively but not simultaneously through the existing catheter, which must therefore be used in exclusive mode by the executing task. *Multiple instances of a task* might be required to unfold repeatable tasks such as therapy or follow-up tasks, although, in many cases, their multiplicity level is unknown at the time of design and can only be determined at the time of execution (either by the user or by the process logic) as it depends on the clinical evolution of a patient or the adequacy of the prescribed treatment. *Delays between iterations of looping activities*, executed as long as a loop condition evaluates to *true*, allow for the expression of things such as "Vancomycin every 8 hours during 7 days" [7]. Furthermore, the *integration of clinical* data is useful for reviewing the availability of process-required data, their location, or the detection of inconsistencies or potential improvements. *Data* are used for input or output to make decisions, initiate a process, coordinate tasks, or inform process actors, whereas *data stores* allow for data persistence beyond the execution of a process model. In addition, in infection CPs, a close relationship between EHR information and model metadata is required to be able to dynamically adapt to clinical happenings. Finally, *overrides at execution time*, either by the user or the system's logic, are required to deal with unexpected happenings during the course of an infection, which can unleash relevant changes in the care process.

### Step B: Identification of Corresponding TP Features for Infection CPs

The TP specification is natively oriented toward the conceptual modeling of clinical processes, thus theoretically providing out-of-the-box capabilities that are present in BPMN only after successive extensions. In this section, we analyze how TP supports the *key features* identified in the previous section. First, *well-structuredness* is a *built-in* feature in the TP standard semantics, which impose implicit restrictions on a process model. *Process modularity* is achieved in TP by the intrinsic grouping of reusable pieces of work in work plans, task plans, and, optionally, subplans for fine-grained work details. TP *events* regulate the temporal behavior of TP tasks or task groups by means of *task-waits* that prevent task execution until facts occur that satisfy a wait condition, such as the occurrence of a task transition or changes in global tracked variables. *Parallel execution of tasks* can be achieved in TP using the *execution type* attribute of the *plan item* top-level class. In addition, *conditional decision* structures exist to represent *if-elseif-else*, *case*, or *event-based* conditions, and the optional *concurrency-mode* attribute defines 4 possible states of a task group during and after the execution of its conditional paths. However, *task duration*, either as constraint or merely informative, does not exist in the current version of TP. When duration is meant as a constraint, TP implements this behavior through *task-waits* associated with events instead. *Relative time constraints between tasks* are managed in TP through *task-waits* for both deterministic and nondeterministic events combined with the use of *dispatchable* tasks. *Resources* are minimally represented in TP through *Resource_Participation* objects required for task execution. The allocation and tracking of workers to a task or task group is done at the time of execution as part of the *materialized* model (ie, the *M** classes, which are only minimally specified in the current TP version). *Multiple instances of a TP task* can be dynamically generated at the time of execution if they are marked as *repeatable* in the process model by the optional *repeat-spec* attribute, which specifies the minimum and maximum number of iterations, and an optional *terminate condition* to exit the *repeat* loop. Each iteration is unrolled into literal sequential copies in the *materialized* image of the work plan. *Delays between iterations of looping tasks* or *task groups* are possible with the optional *period* attribute of a *repeat-spec* associated with a repeatable task or task group to be able to express things such as *daily every 8 hours*, acting as a spacer between the execution of the literal copies of each iteration. Furthermore, *clinical data integration* is indirectly achieved through the *capture data set* attribute, a data set template or form via which data can be introduced during task execution. As TP is part of the openEHR standard, data references are encapsulated using *archetypes*, so the *template-id* attribute specifies the *archetype human-readable identifier* (HRID) to be used, triggering the population of a data set [64]. Finally, *overrides at execution time* are *built-in* in TP, which by default assumes that, at the time of execution, users can have information that is unknown to the system for any number of reasons. The designed process model is performed in an advisory, adaptive way, allowing for logical deletion and addition of tasks at runtime and override of plan parameters such as task execution time or preconditions [64]. Table 1 shows the list of the identified key features and their coverage in extended BPMN and TP.

**Table 1.** Coverage of key features in the extended business process model and notation (BPMN) and Task Planning (TP).

| Feature | Extended BPMN | TP |
|---|---|---|
| Structured workflow definition | SESE[a] restrictions | Built-in |
| Process modularity | Call activity | Dispatchable tasks or task hierarchy or subplans |
| Events | Timer or signaling events | Specialized task transition and state trigger events, among others |
| Parallel execution | Gateways | Execution type and concurrency mode |
| Task duration | BPMN extension | N/A[b] |
| Relative time constraints between tasks | BPMN extension | Built-in task-waits |
| Use of resources | Minimally defined | Minimally defined |
| Multiple tasks | Multiplicity marker | Repeatable tasks |
| Delays between task iterations | N/A | Repeat attribute “period” |
| Data integration | N/A, uses UML[c] | Capture data sets or subject’s proxy services |
| Overrides at the time of execution | Only add activities and events | Built-in (remove or add tasks, plan parameters, or subject preconditions) |

[a] SESE: single entry, single exit.

[b] N/A: not applicable.

[c] UML: Unified Modeling Language.

## Step C: CR-BSI Case Study

### *Typical Process Patterns in Infection Treatment*

The treatment of infections shares some common distinctive patterns that are typical of diseases caused by bacterial microorganisms. First, all types of infections require laboratory tests to determine the pathogen causing the infection in the first place to decide on the most effective antibiotic treatment. In addition, infections must be treated readily to avoid complications that could become life-threatening, following empirical evidence laid out in CGs even though laboratory results are not yet known. Furthermore, antibiotic treatment must follow strict administration rules to be effective, such as dosage, frequency, and duration. Finally, patients must be monitored throughout the entire process to quickly adjust the treatment in case problems arise. We represented some of these distinctive patterns of CR-BSI in TP to empirically check the suitability of TP for modeling infection CPs. The CR-BSI process patterns were originally elicited from the *JHH Antibiotic Guidelines* [15]. The *BPMN-based* CR-BSI process models [43,44] contain both standard BPMN flow objects and extended elements such as *task* and *edge duration*. For modeling the CR-BSI TP flows, we used the open-source draw.io tool, importing the TP-VML libraries [65]. Table 2 shows the correspondence of the extended BPMN process models from the studies by Zerbato [43] and Zerbato et al [44] with the TP process models depicted in this study.

**Table 2.** Correspondence between extended business process model and notation (BPMN) and Task Planning process models of catheter-related bloodstream infection.

| Pattern | BPMN figure number | Description |
|---|---|---|
| 1 | 5 [44] | ET[a] (generic) |
| 2 | None | Determination of ET |
| 3 | 3 [44] | *Staphylococcus aureus* treatment |
| 4 | 4.1 [43] | CLT[b] in coagulase-negative *Staphylococcus* |

[a] ET: empiric treatment.

[b] CLT: catheter lock therapy.

### *Process Patterns in CR-BSI Treatment*

We modeled the process patterns of CR-BSI (Textbox 1) using TP, maintaining the same logic and assumptions made in the study by Zerbato et al [44] when interpreting the JHH CGs.

**Textbox 1.** Process patterns of catheter-related bloodstream infection (CR-BSI).

**Process patterns**

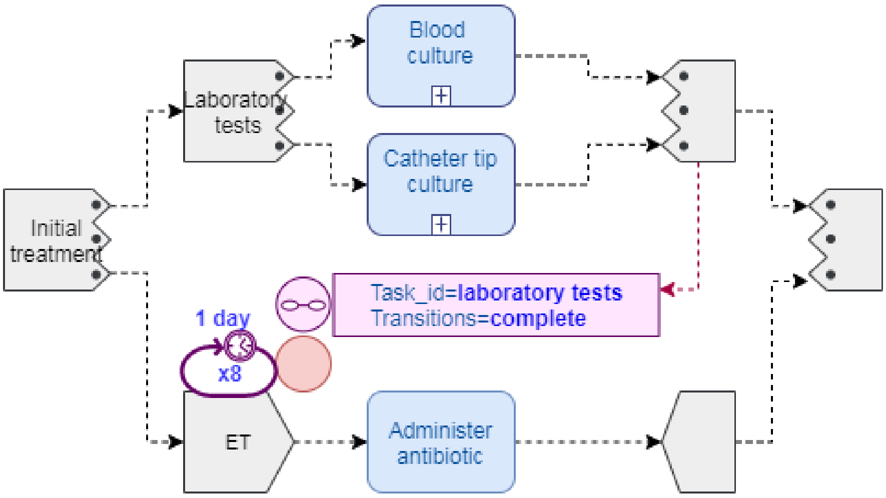
- *Empiric treatment* (ET): when suspicion of an infection exists, the clinical approach is to immediately start treating the patient with a wide-spectrum antibiotic following the empirically gained knowledge laid out in clinical guidelines (CGs) while laboratory tests are ordered in parallel to identify the causing microorganism. As soon as laboratory tests are available, the ET is revised and adjusted if required, either by changing its dosage or duration or replacing it with an organism-specific antibiotic for the concrete pathogen causing the infection.
- *Determination of ET*: this is typically a cognitive decision task that allows clinicians to select the best course of treatment based on CGs or clinician knowledge, clinical facts, and available culture results, if any.
- *Staphylococcus aureus treatment adjustment*: infection treatment usually involves strict management of temporal constraints as antibiotics need a specific administration pace and duration to be effective. During infection treatment, patient evolution and vital signs are continuously monitored to quickly adjust the originally prescribed therapy (eg, its dosage or duration).
- *Catheter lock therapy* (CLT): many infections are caused by the use of catheter implants in hospitalized patients, of which CR-BSI or *urinary tract infections* are typical examples. Intravascular catheters routinely develop microbial communities (biofilms) upon contact with environmental or skin pathogens. When such infections occur and the catheter cannot be removed as it can be counterproductive, antibiotic locks are recommended by CGs as complementary therapy so that both the systemic infection caused by the catheter and the area around the catheter are treated. This means that the *systemic therapy* must be repeatedly but not simultaneously applied with the CLT during a certain period. Both therapies use the same catheter as the vehicle for treatment, so the catheter becomes a shared resource that must be locked exclusively by each therapy task to prevent simultaneous use.

### Pattern 1: Empiric Treatment

The empiric treatment pattern (Figure 2) illustrates the relevance of repeat loops and nondeterministic events in infection CPs. We represented a TP parallel *AND-all-paths* task group with 2 distinct branches: branch 1 represents the ordered set of *laboratory tests*, whereas branch 2 represents the *ET* administration. As more than one laboratory test is ordered, a second inner *AND-all-paths* task group is modeled to indicate that all cultures need to be completed to be able to interpret the results. Once they become available, a nondeterministic *task transition event* is fired, signaling the *ET* repeatable task, which has the *laboratory tests* transition event to the *complete* state as *repeat terminate condition*, ultimately interrupting the *ET* task.

This pattern was represented in extended BPMN [44] using deterministic events instead.

**Figure 2.** Catheter-related bloodstream infection empiric treatment (ET) pattern 1 using Task Planning Visual Modeling Language.



### Pattern 2: Determination of Empiric Treatment

This pattern shows the relevance of encapsulating decisions in rules, which was not possible in BPMN. In case of CR-BSI being suspected, the determination of the empiric treatment requires gathering epidemiology information from the hospital. If *methicillin-resistant* bacteria have high prevalence, a wide-spectrum antibiotic (*Vancomycin*) is usually recommended. Otherwise, a decision between different antibiotic treatments is taken (empiric treatment 1-4 for simplicity) based on the suspicion of a pathogen and on patient conditions. Figure 3 shows a representation using a *gate* task group for the binary decision *high–methicillin-resistant bacteria present*, which, if positive, includes a nested *case* task group with multiple exclusive treatment choices and an inner second binary *gate* to decide between *empiric treatment 2* and *empiric treatment 3*.

However, this pattern can be easily representable via a decision table, so we would rather model it in TP using a *DL* rule evoked from a *determine ET* task, which would encapsulate the knowledge associated with this decision in a single rule to be further maintained and evolved by knowledge experts, allowing for the separation of concerns (Figure 4).

**Figure 3.** Catheter-related bloodstream infection determination of empiric treatment (ET) pattern 2 using Task Planning Visual Modeling Language.

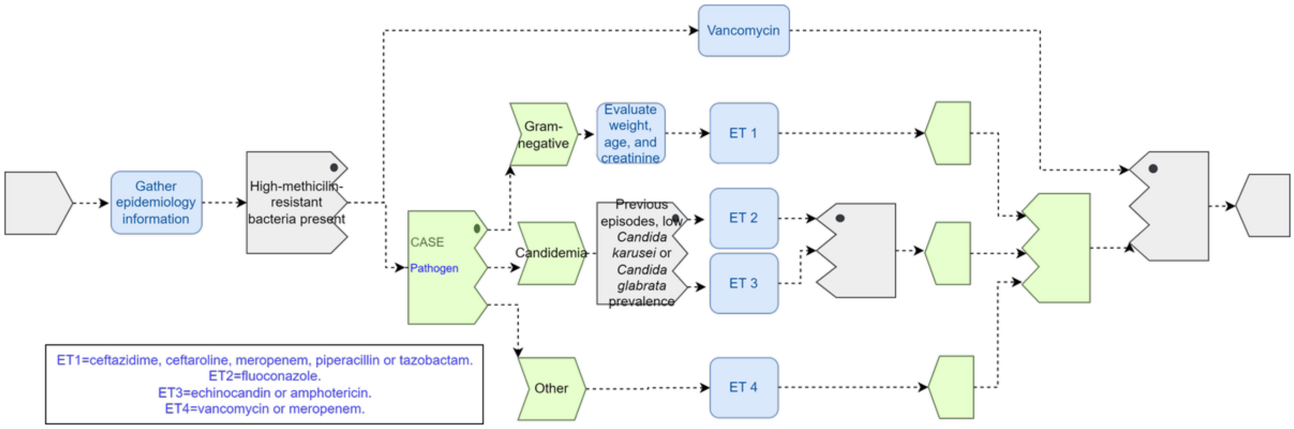
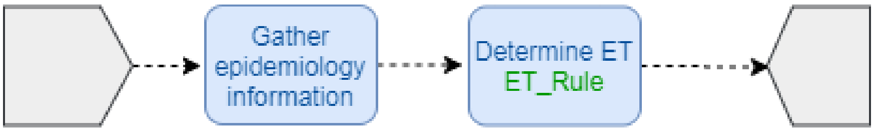


**Figure 4.** Catheter-related bloodstream infection determination of empiric treatment (ET) pattern 2 using Task Planning Visual Modeling Language with a Decision Language rule.



### *Pattern 3: Staphylococcus aureus Treatment Adjustment*

This pattern showcases the importance of overrides in dealing with uncertainty during the course of an infection treatment. The JHH CGs state the following: “Criteria for a 14-day course of therapy: patient is clinically stable; follow-up blood cultures drawn 2-4 days after the initial cultures are negative for *S. aureus*; the patient defervesces with 72 hours of initiation of effective antistaphylococcal therapy. All other patients should receive 4-6 weeks of therapy based on extent of infection” [15].

Figure 5 shows a possible TP model using deterministic events, as in the study by Zerbato et al [44]. The decision *review duration* is a subplan that decides whether therapy duration should continue as planned or be prolonged to a minimum of 28 days and a maximum of 42 days. We used a parallel *and-all-paths* task group with 2 branches: the first branch is a parallel task group *follow-up* representing the *measure temperature* and *check culture results* timeline-driven tasks executed at day 3 and 4, respectively, after treatment start and the second branch is a repeatable task group that executes the initial 14-day *short therapy*. Once the *follow-up* branch ends, the *review duration* subplan is launched asynchronously to determine if the *short therapy* should be prolonged. When the *short therapy* ends, and only if a longer therapy was decided, a *therapy extension* repeatable task is launched until the minimum number of iterations (14) is reached. From that moment on, therapy can be interrupted through a *repeat-spec terminate condition* if a *state trigger* event signals positive patient evolution. The *therapy extension* cycle will end either when patient evolution is positive or, in the most extreme scenario, when the *repeat.upper* limit (28 in our case) is reached. Although this is a good example of how repeatable tasks are of great value for the representation of the course of treatment of infections, we had to perform some calculations with the days to faithfully represent the CPs. For simplicity, we assumed that the treatment was applied once daily but, in reality, antibiotics must be taken in strict time intervals expressed in hours rather than days. Alternatively, a DL rule can be introduced to assess patient evolution as part of the decision logic that determines either a short or a long therapy.

The previous deterministic scenario could be improved, as shown in Figure 6. The main difference is that, in case treatment duration is prolonged by the *adjust treatment* task, a *capture data set* (form) would collect the new treatment information, update the EHR *subject’s proxy* relevant variables accordingly, and trigger a new repeat *override-condition* that would reset the repeat attributes to *repeat.lower=28* (instead of 14) and *repeat.upper=42*. However, this behavior has several implications that will be discussed in the following section.

**Figure 5.** Catheter-related bloodstream infection *Staphylococcus aureus* (*S. aureus*) treatment adjustment pattern 3 (P3) using Task Planning (TP) Visual Modeling Language. JHH: Johns Hopkins Hospital.

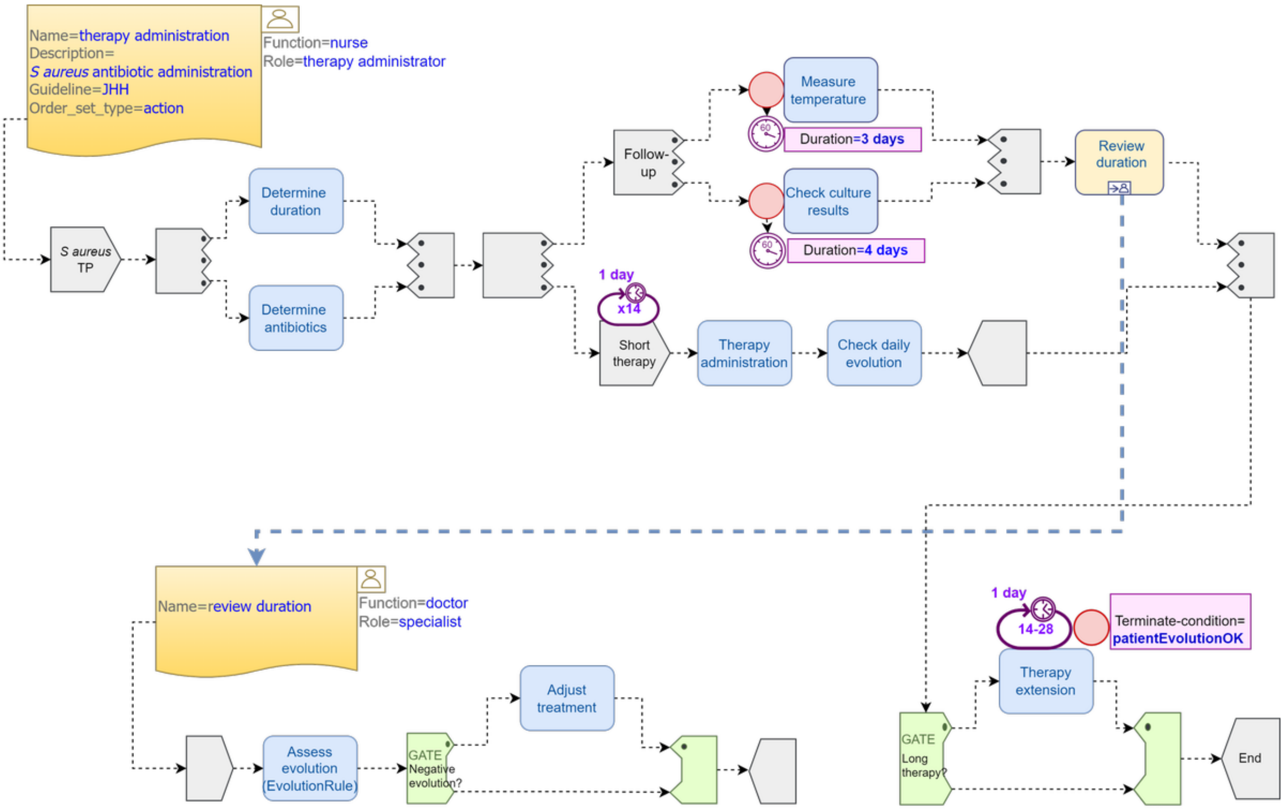
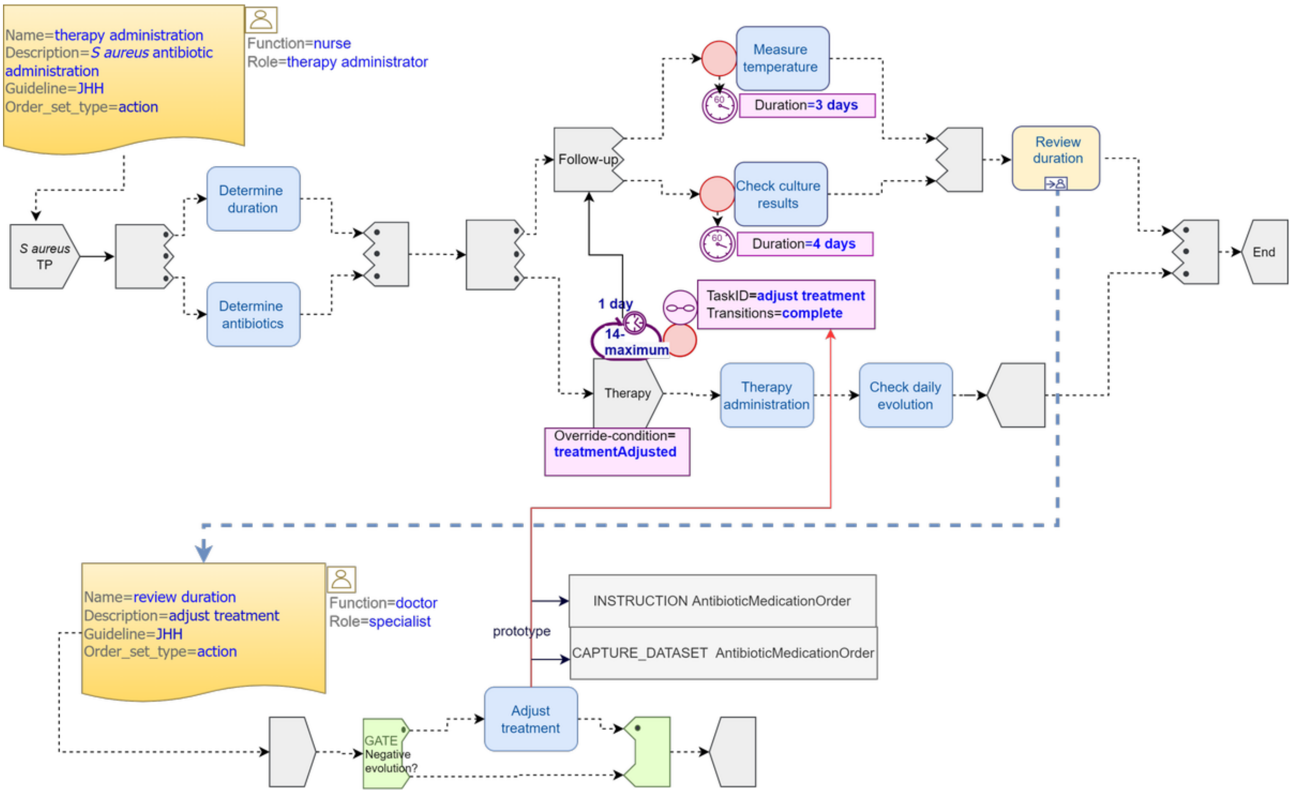


**Figure 6.** Catheter-related bloodstream infection *Staphylococcus aureus* (*S. aureus*) treatment adjustment pattern 3* using extended Task Planning (TP) Visual Modeling Language. JHH: Johns Hopkins Hospital.



### *Pattern 4: CLT*

This pattern illustrates task synchronization when sharing a common resource in exclusive mode. CLT is a technique meant to reduce treatment failure that fills up a catheter with an antimicrobial agent and lets it dwell for a long-enough period. CLT must be administered through the same catheter used for ST, so task synchronization is essential to guarantee catheter availability. Owing to the lack of details in the JHH CGs regarding the alternation and frequency of CLT, we made a few assumptions: (1) both therapies are repeated within fixed but possibly different periods, (2) their timing is independent from each other, and (3) they take place at random but predictable times. With these premises, we modeled 2 repeatable parallel task groups: one including the *CLT* task and the other including the *ST* task. Both task groups first execute an instrumental *“catheter lock”* plan to issue an exclusive hold on the catheter as soon as it becomes available and end with a *catheter release*. The *catheter lock* includes a *repeatable* task group that checks catheter availability and ends as soon as the catheter becomes available, moment at which the catheter is locked, and a *callback notification* is sent to the calling task (either *ST* or *CLT*). Once the corresponding therapy task is executed, the trailing *catheter release* task is called to release the catheter (Figure 7).

Ideally, this lock mechanism should be a *built-in* behavior in TP when a resource is defined as *exclusive*, greatly simplifying the process model as both the instrumental *catheter lock* and the *catheter release* task plans would not be necessary (Figure 8).

**Figure 7.** Catheter-related bloodstream infection catheter lock therapy (CLT) pattern 4 using Task Planning Visual Modeling Language. CG: clinical guideline; JHH: Johns Hopkins Hospital; ST: systemic therapy.

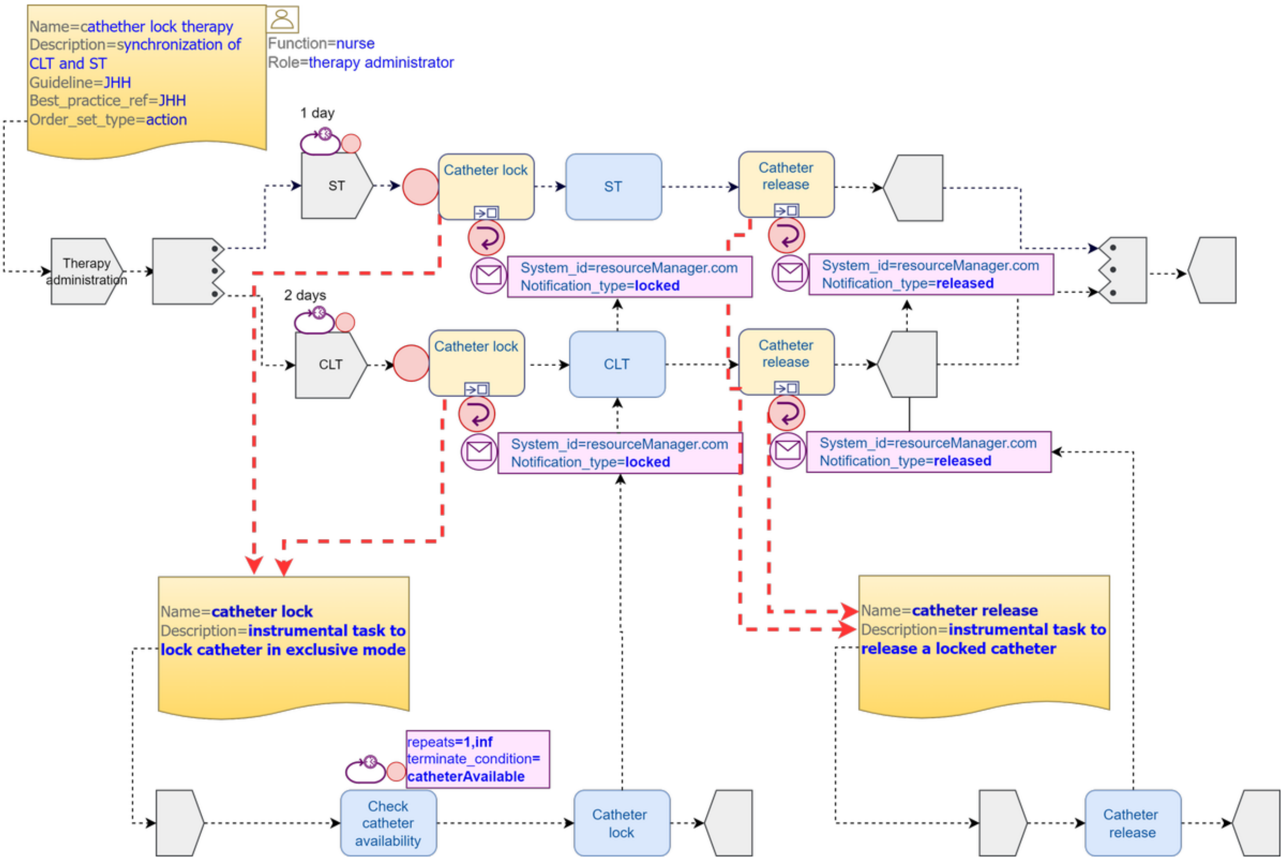
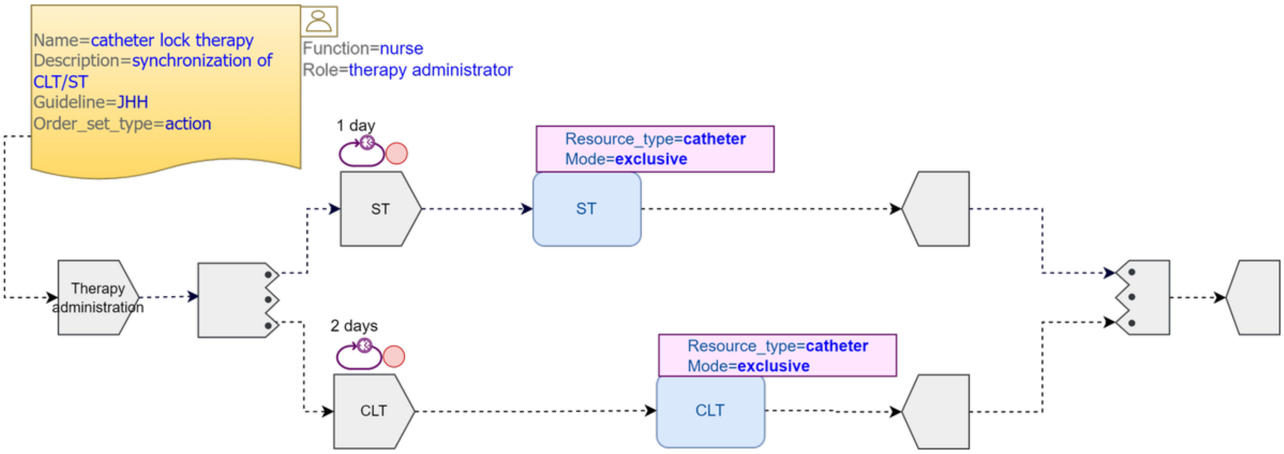


**Figure 8.** Catheter-related bloodstream infection catheter lock therapy (CLT) pattern 4* using extended Task Planning Visual Modeling Language. JHH: Johns Hopkins Hospital; ST: systemic therapy.



## Step D: Analysis and Possible Extensions

### *Overview*

We were able to model typical process patterns of CR-BSI in TP (Figures 2-8) keeping the logic applied to the interpretation of the JHH CGs in the study by Zerbato et al [44] with no amendments or extensions to the current version of the TP standard. According to the 4 requirements for CP languages [58], the TP specification fulfills *requirement 1* with the openEHR architecture and foundation classes as well as many of the requirements *2*, *3*, and *4* by means of parallel concurrency modes extensible with specific rules, *dispatchable* and *performable tasks* associated with *defined actions*, or temporal constraints between tasks using advanced events. In this study, we in fact analyzed the suitability of domain-specific features of *requirements 2 to 4* for infection CPs. However, in some cases, we depicted alternative process models using a different approach (eg, pattern 3* and pattern 4*) by proposing extensions that enhance the TP standard capabilities for modeling infection CPs, such as new annotation capabilities of tasks, enhanced cyclic constructs, new override behavior of the TP engine, new events, and enhanced resource management. To determine these potential TP extensions, we followed the methodology proposed in the study by Braun and Schlieter [58] for BPMN extensions, which we consider applicable for TP, consisting of 6 steps that extend the approach of Stroppi et al [66], resulting in an *equivalence* or *no equivalence* conclusion. A *no equivalence* conclusion can have 3 reasons: (1) the entire concept is missing (*extension concept*), (2) a relationship between 2 concepts is missing (*association between concepts*), or (3) attributes owned by a concept are missing (*property of a concept*).

Textbox 2 shows a summary of the main TP constructs used in the CR-BSI models. Models where an extended approach was proposed are denoted by a trailing *.

In the following sections, we analyze the most frequently used TP constructs and their possible improvements, which could lead to potential extensions of the standard.

**Textbox 2.** Main Task Planning (TP) constructs used in each use case.

**TP construct and patterns in which it was used**

- And-all-paths parallel task group (TG): pattern 1, pattern 3, pattern 3*, pattern 4, and pattern 4*
- Repeat task or TG: pattern 1, pattern 3, pattern 3*, pattern 4, and pattern 4*
- Gate binary decision group: pattern 2, pattern 3, and pattern 3*
- Timeline event: pattern 3 and pattern 3*
- Asynchronous dispatchable task: pattern 3 and pattern 3*
- Task transition event: pattern 1 and pattern 3*
- Synchronous dispatchable task: pattern 4
- State trigger event: pattern 3
- System notification event: pattern 4
- Case decision group: pattern 2
- Decision Language rule: pattern 2*

### *Parallel Execution and Decision-making Constructs*

The 4 TP *concurrency modes* associated with the parallel *execution type* cover a wide range of situations that might arise in actual clinical scenarios and were expressive enough for the representation of CR-BSI patterns. The *and-all-paths concurrency mode* was most frequently used, followed closely by the *gate decision group* construct, for single binary decisions. Occasionally, we used a *case decision group* in pattern 2, a cognitive diagnosis decision-making task that can be better represented as a single DL rule. Throughout the modeling process, we did not detect any need for improvements in the existing TP decision constructs or in parallel execution. However, the standard allows for the definition of rules if more sophisticated parallel behavior is required in complex scenarios.

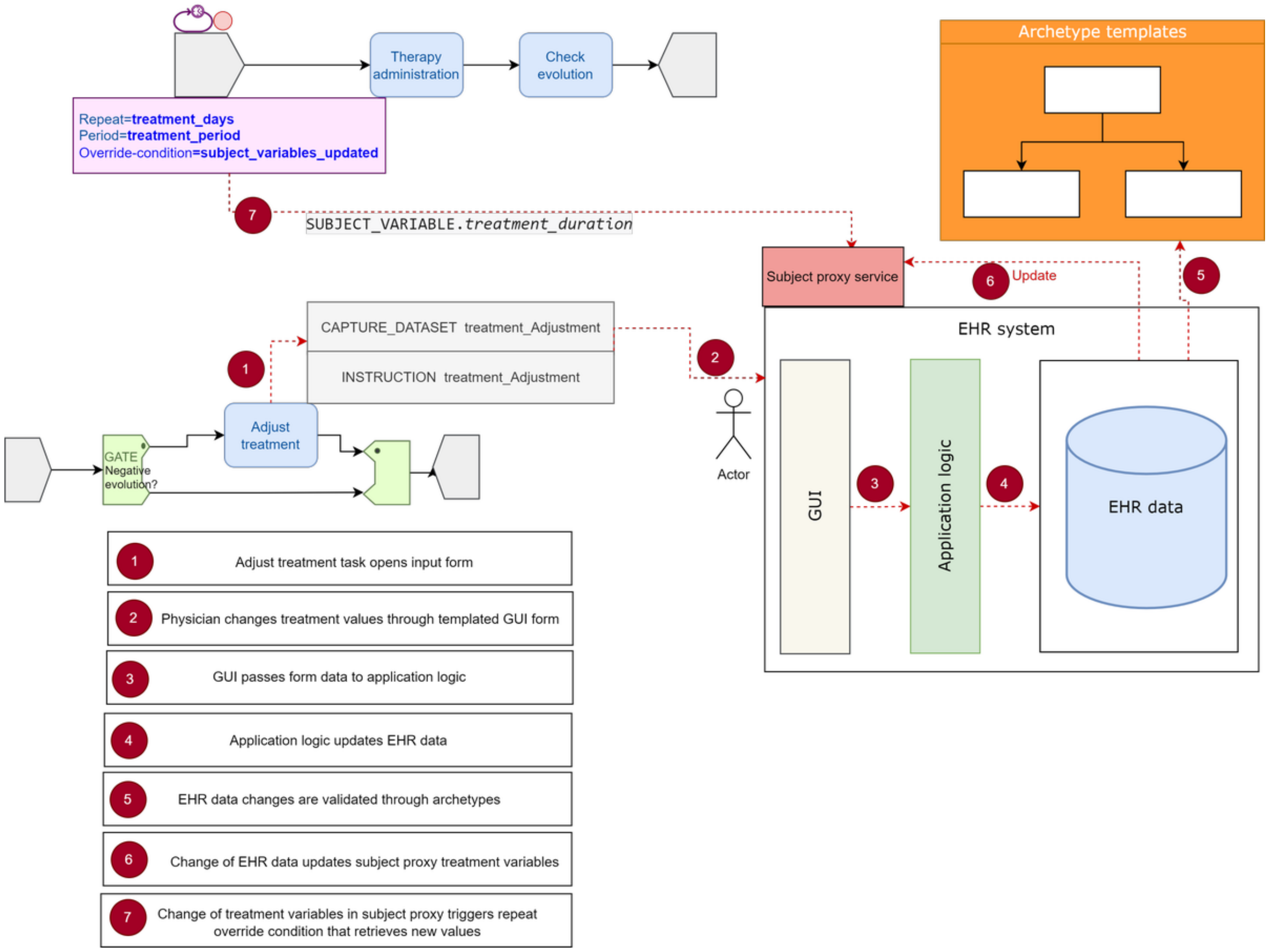
### *Repeatable Task Constructs*

The second most used TP construct was the task or task group *repeat*, widely used for medication plans, therapy administration, or monitoring of patients' vital signs. Antibiotic treatments usually have the form *Ciprofloxacin 400 mg IV Q8H*, which can be represented in TP using the *repeat* attribute of a *plan item*, which unfolds copies of the successive antibiotic intakes in the TP execution engine. This attribute specifies a minimum and, optionally, a maximum number of iterations, a time interval or *period* between iterations, and a *terminate condition* evaluated as soon as the minimum number of iterations is reached before the execution of new iterations. The *terminate condition* can be any *plan definition* event exiting the *repeat* loop whenever event conditions are met. The TP *repeat* construct is a hybrid loop as it mimics both a *for* programming loop until the *repeat.lower* limit is reached and, after that, a *while* programming loop. This construct has proven to be essential for modeling CR-BSI CPs, which is why it could be further enhanced to achieve its full potential. For example, in pattern 3, treatment duration is adjusted *on the fly* when the patient does not evolve as expected—the treatment starts as a 14-day–long repeatable *short therapy* task and transitions smoothly to a 28- to 42-day *therapy extension*. The model is improved in the alternative pattern 3* to avoid the need to represent 2 distinct therapy tasks by proposing a mechanism to signal a repeatable task not to terminate but to modify its attributes while still running; that is, an override of the TP metadata model triggered by the process intrinsic logic. The main advantage of pattern 3* is a much simpler visual representation and a more powerful *repeat* construct with adaptive capabilities to respond to events not known at the time of process design. However, this behavior assumes a few things: first, the patient's medication record is stored in the EHR and available to the TP engine; second, whenever treatment is adjusted, it first updates the EHR through a *capture data set*, generating a new medication instruction; third, there is some predefined mapping mechanism between the medication instruction and the *repeat* metadata, used by the TP engine upon the execution of the *adjust treatment* task (eg, taking the information in the *ORDER REF* class, which "represents a logical tracking reference to one 'order' in the real world" [14]); finally, the pattern 3* model would also require modifications in the TP abstract syntax, a new repeat *override condition* attribute in the *Task_Repeat* class to trigger the reset of the repeat attributes using the new treatment information in the *subject's proxy* (Figure 9), and changes in the TP engine to keep internal state records of repeat iterations and dynamically update the materialized literal copies, unfolding pending iterations and considering the ones already consumed.

The *Task_Repeat* attributes would be automatically bound to the patient's EHR medication record through a *defined action* associated with the task and its corresponding *prototype* based on its turn on a *medication archetype* containing a structured description of the therapy administration and time. When the *adjust treatment* task is performed, an *action* instance would be created from the *prototype* reflecting any divergences from the planned form of the task. To support this scenario, we suggested the extension of the *Task_Repeat* class, as outlined in Table 3.

The execution logic of our extended *Task_Repeat* class would be as follows: if *start-condition* is true and *repeat.lower* is >0, a loop is started of *1* to *repeat.lower* iterations (the *for* part of the *repeat* loop). This minimum number of iterations is executed unconditionally unless a *skip-condition* is met in any of the iterations. Only when iteration number *repeat.lower+1* is reached the rest of the conditions could apply; that is, an *override* or *terminate condition*.

**Figure 9.** Possible mechanism of *Staphylococcus aureus* treatment adjustment pattern 3* with repeat override condition. EHR: electronic health record; GUI: graphical user interface.



1. Adjust treatment task opens input form
2. Physician changes treatment values through templated GUI form
3. GUI passes form data to application logic
4. Application logic updates EHR data
5. EHR data changes are validated through archetypes
6. Change of EHR data updates subject proxy treatment variables
7. Change of treatment variables in subject proxy triggers repeat override condition that retrieves new values

**Table 3.** Proposed new attributes of Task Planning classes.

| Class and attribute | Explanation | Possible values |
|---|---|---|
| **Task_Repeat** | | |
| Terminate-condition | Add more elaborated expressions or rules, possibly using DL[a] | N/A[b] |
| Start-condition | New condition, expressed in a similar manner as the terminate-condition, would allow the first iteration to occur only if both the minimum iteration is not 0 and this condition evaluates to true | N/A |
| Skip-condition | New condition, expressed in a similar manner as the terminate-condition, would skip the currently executing iteration if it evaluates to true by eliminating the materialized literal copy both in the "for" and the "while" parts of the loop. | N/A |
| Override-condition | New condition, expressed in a similar manner as the terminate-condition, would allow for the overriding of Task_Repeat metadata at the time of execution | N/A |
| **Resource_Participation** | | |
| Category | Categorize resources with clinical process relevance | • External or implant<br>• Medical device |
| Mode | Exclusive or shared mode | • Exclusive or shared |
| State | State of clinically relevant resource | • Allocated or locked<br>• Deallocated or unlocked |
| **Plan_Item** | | |
| minDuration | Minimum duration of Plan_Item | N/A |
| maxDuration | Maximum duration of Plan_Item | N/A |
| timeUnits | Units of time to express duration values | N/A |
| durationUse | Specifies if the Plan_Item duration is information or also constraint | • Informative or prescriptive |

[a] DL: Decision Language.

[b] N/A: not applicable.

### Resource Constructs

In infections, catheters are used as a vehicle for the administration of different therapies sometimes incompatible with each other, as we have seen in pattern 4 with *CLT* and *ST*, where we had to define 2 instrumental *catheter lock* and *catheter release* tasks, unnecessarily increasing flow complexity. In the alternative pattern 4* model, we eliminated the need for them by extending the *resource participation* class with new attributes to represent clinically relevant resources, the state of which can change at the time of execution owing to unexpected clinical happenings. This would require a *resource state machine* in the TP *materialized* model, similar to the TP engine *task state machine*, that would help improve the *resource perspective* of the TP standard, similar to how it was done for BPMN [59]. Accordingly, the TP engine should adopt a new specific behavior for resource allocation or deallocation at the time of execution. In the pattern 4* model, the *catheter* resource could be tagged as of *exclusive mode* so that the TP engine can exclusively allocate the resource to the executing task, implementing a *built-in* exclusive lock and release strategy to avoid *deadlocks*. Table 3 shows the proposed new attributes of the *Resource_Participation* class.

The proposed extensions affect the abstract syntax; that is, the TP standard metadata model and the TP engine logic. In addition, the EHR could document resources used in the patients' medical procedures, especially those implanted that can cause or aggravate an infectious disease. For example, in CR-BSI or urinary tract infection, it is relevant to know if the catheter has been removed, irrigated, replaced, or salvaged to determine treatment duration.

### Events

The TP *task-waits* allow for the introduction of dynamically determined delays between tasks, adding fundamental modeling capabilities through the use of specialized, *nondeterministic* events: *task transition*, *state trigger*, *callback*, *manual*, and *system notification*. The *task transition event* makes use of the TP engine *state machine* representing all possible states and transitions between tasks, which is essential for dynamically detecting task completion in infection treatment processes. The *state trigger event* signals changes in the values of global monitored variables representing patients' vital signs. Together with the set of TP *deterministic events*, they provide advanced modeling capabilities in TP, greatly optimizing workflow execution times. We made use of a *task transition event* in pattern 1 or a *state trigger event* in pattern 3, which is consistent with the uncertainty and time-driven nature of the course of infection treatments. To complement these *nondeterministic* capabilities, we suggest the addition of a new *resource state event* to improve the *resource perspective* of TP, signaling tasks when resource allocation, deallocation, or exclusive locks or releases occur through a new *resource state machine* in the TP engine.

### *Temporal Constraints Between Tasks*

The treatment of infections requires a rather strict regime of temporal constraints to optimize its effectiveness and, at the same time, limit antibiotic intake to the strictly necessary levels to avoid the development of antibiotic resistance. The BPMN *duration* extension, widely used in the study by Zerbato et al [44], does not currently exist in TP, and it has been replaced in our models with dynamic *nondeterministic* events, as in pattern 1, where a *task transition event* triggers a *repeat terminate-condition* to interrupt the *empiric treatment* loop. A *nondeterministic* event-driven model can speed up a treatment process (eg, when laboratory tests are ready earlier than expected, the pathogen-specific treatment can be started). Despite not being strictly required in our TP models, there could be other thinkable scenarios in which task *duration* could be useful, either as a constraint owing to regulations or in high-precision treatments. Arguably, medication records with duration information are already documented in the EHR; thus, a task *duration* attribute in the TP meta model could be seen as redundant. However, we think that a workflow task should have an intrinsic duration inherent to any piece of work which can be relevant for prescriptive tasks and overrides at the time of execution (eg, in a treatment adjustment) to dynamically change the number of antibiotic intakes. Consequently, we suggest adding optional *duration* attributes to the TP *Plan_Item* parent class (Table 3). The new attributes should be able to be overridden on demand at the time of execution and leave the corresponding accounting records for both audit and process-mining purposes. In addition, the concrete syntax (ie, the TP-VML graphical representation) could be enhanced with new annotation capabilities.

Furthermore, TP uses different *waits* to hold task execution: *task*-, *timer*-, and *callback-waits*. A list of events to be *waited on* can be specified in the *events-list* attribute of the *task-wait* class, both deterministic and nondeterministic. The *task-wait* ceases when any of the events in the *events-list* becomes true and is currently evaluated as a logical OR (cursive). We propose to add an attribute *event-list-relation* in the *task-wait* class with possible values *OR* and *AND* so that a task can wait on the simultaneous occurrence of more than one event, a typical behavior of *complex event processing* systems. However, this requires other possibly deep-going adaptations of the TP logic to evaluate the occurrence of multiple events within a given time window, so a thorough cost-benefit analysis should be performed on this specific feature. Furthermore, TP has an optional *event-relation* attribute in the *Task-wait* class for increased time granularity, intended to "allow a task to be specified as commencing before, with or after the triggering event (such as a meal)" [14]. In the case of deterministic events, this attribute could be further refined with a new *offset* attribute expressed in positive or negative time units relative to the triggering event as a *time quantifier*. As a *task-wait* can wait for more than one event, we would interpret the newly quantified event-relation as *transition task* t1 *to* the available *state* x *hours* before|after *the* events-list *via which* t1 task-wait *is ceased* in *AND* or *OR* logic; that is, all triggered events (*AND*) or just 1 (*OR*). The effect of this is to add or subtract a concrete amount of time (the *offset x*) to the firing time of the triggering event or events associated with the *task-wait*, causing the *task* to either be delayed or advanced. In the case of *nondeterministic* events, the *event-relation* attribute cannot have quantifiers, at least for the *before* operand, as the time of occurrence of such events is unknown, so the task can only become available immediately *before* or *after* the set of triggering events.

We have more generally analyzed TP time dependencies between tasks from a pragmatic point of view, checking if they meet the 13 Allen time interval operands [67], as summarized in Table 4. We focused on *performable tasks* as *dispatchable tasks* have their own synchronization mechanism through specific waits and events. We found that three pairs of the 6 symmetric Allen interval operands on task execution are naturally implemented in the TP standard: (1) *precedes* or *preceded by* is built upon the default behavior of the TP execution type *sequential*, which allows a task *t2* to become available when the predecessor task *t1* is completed plus an additional *task-wait* in the successor task *t2*; (2) *meets* or *met by* is a generalization of the previous case also built upon the execution type *sequential*, which allows a task *t2* to become immediately available for execution when the predecessor task *t1* is completed provided that no *task-waits* are included (this is the default behavior of *sequential* tasks); and (3) *starts* or *started by* is a temporal behavior achieved by modeling a parallel task group with tasks *t1* and *t2* and no *task-waits* included, so tasks *t1* and *t2* start at the same time.

The other 3 pairs of symmetric Allen interval operands plus the *equals* operand impose restrictions upon the tasks' end or start times, always in parallel execution settings, and need more fine-grained time expressions, which is not always easy to achieve in the current version of the TP standard (Textbox 3).

Table 5 shows a list of the proposed TP extensions for modeling infection CPs, which we believe could also be of interest in other complex clinical settings. Most of the suggested extensions are labeled as either "property of a concept," enriching already existing concepts, or as "extension concept" [58], mostly referring to a new behavior of the TP engine or materialized classes model.

Finally, Table 6 shows a comparison of the initially identified *key features* in extended BPMN, standard TP, and our proposed *extended TP* (represented in the *TP** column) providing an overview of the possible areas of enhancement.

**Table 4.** Possible Task Planning (TP) representation of 13 Allen time relation operands.

| Temporal relation | Graphic | Implementation in TP |
|---|---|---|
| Precedes (p) and preceded by (P) | | Sequential execution with “task-wait” or “period” in repeatable tasks |
| Meets (m) and met by (M) | | Sequential execution with no “task-wait” or “period” in repeatable tasks |
| Overlaps (o) and overlapped by (O) | | “Plan-time-origin” extension with new attribute “plan item” |
| Finishes (f) and finished by (F) | | “Plan-item” class with new attribute “end time” or extended “task-wait to complete” |
| Contains (D) and during (d) | | “Task-wait to complete” with triggering event “t2 completed”, plus time quantifier |
| Starts (s) and started by (S) | | Parallel “task group” with no “task-wait” or with the same “task-wait” |
| Equals (e) | | Parallel “and-all-paths” “task-group” with no “task-wait” and same task duration |

**Textbox 3.** The remaining 3 pairs of symmetric Allen interval operands plus the equals operand.

**The remaining operands**

- *Overlaps or overlapped by*: this relation can be modeled using a *parallel task* group (TG) and inserting *task-waits* either on the overlapping or the overlapped task. If the overlapping part of both tasks needs to be specified more accurately, this pattern should either include global timeline-specified *task-waits* for each task relative to the work plan start time or, ideally, time-lined *task-waits* with an offset relative to the first starting task (t1 or t2). However, this last scenario would require an extension of the *Plan_time_origin* enumeration class with a new attribute *plan-item* to specify a list of plan items of the parallel TG to be used as a relative reference for the timeline offset.
- *Finished by* or *finishes*: in this temporal pattern, parallel tasks *t1* and *t2* need to end at the same time. Generally speaking, an *and-all-paths parallel* task group could be used as it ends only when all TG branches are finished. However, the time to completion cannot be enforced as it depends on the longest executing branch, and it does not allow for the limitation of this behavior to 2 branches, for example. Thus, for a compulsory end time for *t1* and *t2*, a deterministic *end time* restriction should be applied to each task of the parallel TG. We could not find a way to implement this in the current Task Planning (TP) version, so a possible solution could require an extension of the *plan item* class with a new attribute *end time*: *ISO 8601 duration* or, alternatively, *task-waits* should apply to transition a task not only to the *available* state but also to any state or, at the very least for this specific scenario, to the *completed* state.
- *Contains* or *during*: this temporal relation states that *t2* must be started after *t1* has started and must end before *t1* ends (or vice versa). The start of *t2* after *t1* can be implemented as in the *overlaps* or *overlapped* case. However, the end of *t2* before *t1* is not that obvious to implement as *task-waits* are not an option because both tasks are already being executed by then. In the current TP version, we could not find a reasonable way to express that t2 *must end* x *units of time before the end of* t1 as a relative time constraint. To make it feasible, *task-waits* should apply to transition a task not only to *available* state but also to any state or, at the very least, to the *completed* state, as in the *finished by* or *finishes* pattern. In that case, a combination of a *task-wait to complete* having as a triggering event the end of *t2* with a time quantifier would allow for a more fine-grained expression of this pattern.
- *Equals*: this temporal constraint states that tasks *t1* and *t2* must start and end at the exact same time. For example, this can be achieved with a parallel TG with no task-waits and a specific restriction on each task duration. Thus, a possible solution would require the *duration* extension proposed previously.

**Table 5.** Summary of proposed Task Planning (TP) extensions for infection clinical pathways.

| Extension | Class | Reason | Origin | Affects | Type |
|---|---|---|---|---|---|
| New “duration” attribute | Plan_Item | Informational, support Allen “equals” pattern | Pattern 1 | AS[a] | PoC[b] |
| New repeat condition attributes | Repeat-spec | Enrich “repeat” construct | Pattern 3* | AS | PoC |
| New repeat “terminate-condition” value | Repeat-spec | Allow DL[c] rule | Pattern 3 | AS | PoC |
| New repeat behavior | TP engine | Override “repeat” metadata at the time of execution | Pattern 3* | L[d] | EC[e] |
| New resource attributes | Resource-participation | Extend resource perspective | Pattern 4* | AS | PoC |
| New resource allocation | TP engine | Implement behavior for exclusive resources | Pattern 4* | AS | EC |
| New resource transition event | Events | Detect resource state transitions | Pattern 4* | AS and CS[f] | EC |
| New resource state machine | TP engine | Specify transition between possible states of resources | Pattern 4* | AS | EC |
| New “event-list-relation” attribute | Task-wait | Allow for the specification of the logical relation between multiple triggering events (OR and AND) | Literature review | AS | PoC |
| New “offset” attribute | Task-wait | Allow for the delay or advancement of a deterministic event | Literature review | AS | PoC |
| New “resume-type” value | Resume-action | Allow for a rule as “resume-type” | Literature review | AS | PoC |
| New “end-time” attribute | Plan_Item | Support Allen “finished by” or “finishes” and “contains” or “during” | Allen | AS | PoC |
| New “start-time” attribute | Plan_Item_Origin | Start time of a plan item as reference for timeline in “overlaps” | Allen | AS | PoC |
| New “task-wait” behavior | TP engine | Task-waits on task transition to completed (“finished by” or “finishes” and “contains” or “during”) | Allen | L | EC |

[a] AS: abstract syntax.

[b] PoC: property of a concept.

[c] DL: Decision Language.

[d] L: logic.

[e] EC: extension concept.

[f] CS: concrete syntax.

**Table 6.** Comparison of key features of extended business process model and notation (BPMN), Task Planning (TP), and extended TP.

| Feature | Extended BPMN | TP | TP* |
|---|---|---|---|
| Structured workflow definition | Extension SESE[a] | Built-in | —[b] |
| Process modularity | Call activity | Dispatchable tasks in synchronous or asynchronous mode or subplans | New “resume-type” in synchronous dispatch |
| Events | Timer and signaling events | Specialized task transition and state trigger events | New “resource transition event” |
| Parallel execution | Gateways | Concurrency mode | — |
| Task duration | BPMN extension | N/A[c] | New task duration attributes |
| Relative time constraints between tasks | BPMN extension | Built-in “task-waits” | New “event-list-relation,” offset, “Plan_Item,” “plan-item-origin” attributes, and “TP engine” behavior |
| Use of resources | Minimally defined | Minimally defined | New resource attributes and “TP engine” behavior |
| Multiple tasks | Multiplicity marker | Repeatable tasks | New repeat conditions |
| Delays between iterations of looping tasks | N/A | Repeat attribute “period” | — |
| Data integration | N/A, uses UML[d] | “Capture data sets” and “subject’s proxy” services | Mapping between EHRs[e] and TP metadata |
| Overrides at the time of execution | Few exceptions (add activities or events) | By design to remove or add tasks, plan parameters, and subject preconditions | On-the-fly override of model metadata (repeat) information |

[a] SESE: single entry, single exit.

[b] None.

[c] N/A: not applicable.

[d] UML: Unified Modeling Language.

[e] EHR: electronic health record.

## Discussion

### Principal Findings

The results of our experiments show, on the one hand, the native suitability of TP for infection management and possibly for other complex clinical scenarios, confirming our initial hypothesis, and, on the other hand, the potential of the TP standard to simplify complex CPs by extending some of its native capabilities following the methodology for BPMN extensions laid out in the study by Braun and Schlieter [58], which we consider applicable to TP. The proposed TP extensions are mainly focused on increased synchronization between tasks, supporting Allen temporal relations, new dynamic behavior of repeat constructs, and enhancements in the resource perspective to emphasize the relevance of medical-associated devices, such as implanted catheters, in the course of infectious diseases. When applicable, we also proposed the use of complementary specifications such as DL.

In pattern 1, we saw how the rich set of nondeterministic TP events is better suited to represent uncertainty about task duration than the deterministic events used in the corresponding BPMN model, allowing for a dynamic response to unknown clinical happenings and, as a result, faster process execution. Nevertheless, we proposed to enhance the TP abstract syntax with new task duration attributes to allow for fine-grained fixed temporal constraints that could be of use in specific scenarios.

In the case of pattern 2, we initially modeled it in the depicted pattern 2 model as a cognitive decision-making task expressed as a multiple decision tree type of task, as done in BPMN. However, we proposed an alternative pattern 2* TP model to encapsulate the complexity of the cognitive decision-making task in a single rule instead. To this end, the TP standard is complemented by the DL specification, allowing for the evocation of DL rules from within a task or task group, which is not possible using BPM alone, resulting in simpler workflows and improved maintenance and evolution of task-related knowledge.

Pattern 3 highlights the need for a more advanced adaptive behavior of the TP execution engine to address new facts or clinical happenings that cannot be foreseen in advance in static BPM models, allowing for dynamic changes in the process logic at the time of execution. The pattern 3 TP model mimics the corresponding BPMN model, with 2 distinct cyclic therapy tasks. However, in the alternative pattern 3* model, we proposed extending TP to enhance the override capabilities of the *repeat* loop, which is widely used in infection CPs to represent, for example, cyclic medication tasks. Dynamic overrides were not available in BPMN, resulting in an unnecessary increase in the workflow complexity.

Finally, pattern 4 shows the synchronization between 2 tasks that cannot be executed simultaneously through the same device; in our case, a catheter. This pattern could not be represented in

BPMN owing to its complexity. We could represent it easily in the pattern 4 TP depicted model using *out-of-the-box* TP constructs by modeling 2 instrumental tasks for synchronization purposes complemented with the rich set of nondeterministic TP events. However, in the alternative pattern 4* model, we showed how enhancing the TP standard with a resource extension could greatly simplify the pattern 4 representation by adding new attributes to the *Resource_Participation* class and proposing both a new nondeterministic *resource_transition* event and a new *state machine* in the TP engine. The proposed TP extension could be generally used to represent medically relevant devices such as catheters, mechanical ventilators, or any other disease-associated devices, thus improving the TP native *resource perspective*.

Furthermore, we theoretically analyzed other possible extensions that could be of use in infection management and other clinical scenarios, such as adding *complex event processing* capabilities to the standard, an enhanced *task-wait* condition logical expression using the newly proposed *event-list-relation* and *offset* attributes, and a new *Resume_type* in a task synchronous dispatch. Finally, we examined the TP temporal restrictions between tasks from a theoretical point of view, analyzing how the TP standard addresses the 13 Allen temporal relations and proposing extensions to support 3 of the 6 symmetric Allen relations plus the *equals* temporal operator. However, the proposed temporal extensions must be weighed against their implementation costs as they require an enhanced *task-wait* logic for a task to transition not only to the *available* state but also to the *completed* state, as well as new *time reference* attributes to be able to express more sophisticated time constraints between tasks.

### Strengths and Limitations

The proposed extensions provide a more consistent approach to dealing with uncertainty, thereby improving overall workflow efficiency and duration by adding dynamic response capabilities during workflow execution. They also acknowledge the importance of clinically relevant resources in the course of infection treatment, improving their visibility and clinical role.

A possible limitation is that many of the proposed extensions require changes in the abstract syntax and the behavior of the TP engine; for example, in pattern 3* to dynamically monitor and reset the *repeat* loop count. These changes in the standard should be weighed against the expected benefits. In addition, the use of rules within CPs to represent knowledge-intensive tasks requires further definition of the DL specification, which is still under development. Furthermore, the use of resources in CPs should be synchronized with the resource information defined in the EHR through specific mechanisms. Finally, the overall suitability of the standard for other complex clinical domains should be further assessed in future studies addressing the specific domain constraints.

### Future Directions

Future work will focus on analyzing the DL specification [50] in relation to TP, a new *openEHR* formalism for evoking rules from TP expressed using the *openEHR Expression Language* [68] and based on the *openEHR Basic Meta Model* [51]. We will also analyze the new OMG *BPM+ Health* initiative aiming to promote the use of the 3 OMG standards—BPMN, Case Management Model and Notation, and Decision Model and Notation—in health care domains [69].

## Acknowledgments

This work was partially funded by the SITSUS project (reference RTI2018-094832-B-I00) and the Trust in Artificial Intelligence for Hospital Antimicrobial Resistance project (CONFAINCE reference PID2021-122194OB-I00) and supported by the Spanish Ministry of Science and Innovation, the Spanish Agency for Research (MCIN/AEI/10.13039/501100011033), and, as appropriate, by the European Regional Development Fund A way to make Europe.

## Conflicts of Interest

None declared.

## References

1. Haug PJ, Narus SP, Bledsoe J, Huff S. Promoting national and international standards to build interoperable clinical applications. AMIA Annu Symp Proc 2018;2018:555-563 [FREE Full text] [Medline: 30815096]
2. Campbell H, Hotchkiss R, Bradshaw N, Porteous M. Integrated care pathways. BMJ 1998 Jan 10;316(7125):133-137 [FREE Full text] [doi: 10.1136/bmj.316.7125.133] [Medline: 9462322]
3. 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]
4. Vanhaecht K, De Witte K, Depreitere R, Sermeus W. Clinical pathway audit tools: a systematic review. J Nurs Manag 2006 Oct;14(7):529-537. [doi: 10.1111/j.1365-2934.2006.00705.x] [Medline: 17004963]
5. 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(1):31 [FREE Full text] [doi: 10.1186/1741-7015-8-31] [Medline: 20507550]
6. 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]
7. Business Process Model and Notation. Object Management Group (OMG). URL: https://www.omg.org/spec/BPMN [accessed 2020-09-26]

8. Combi C, Oliboni B, Gabrieli A. Conceptual modeling of clinical pathways: making data and processes connected. In: Artificial Intelligence in Medicine. Cham: Springer; 2015.
9. 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]
10. Greenes RA, Bates DW, Kawamoto K, Middleton B, Osheroff J, Shahar Y. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. J Biomed Inform 2018 Feb;78:134-143 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.005] [Medline: 29246790]
11. Ramos-Merino M, Álvarez-Sabucedo LM, Santos-Gago JM, Sanz-Valero J. A BPMN based notation for the representation of workflows in hospital protocols. J Med Syst 2018 Aug 29;42(10):181. [doi: 10.1007/s10916-018-1034-2] [Medline: 30155797]
12. Ruiz-Fernández D, Marcos-Jorquera D, Gilart-Iglesias V, Vives-Boix V, Ramírez-Navarro J. Empowerment of patients with hypertension through BPM, IoT and remote sensing. Sensors (Basel) 2017 Oct 04;17(10):2273 [FREE Full text] [doi: 10.3390/s17102273] [Medline: 28976940]
13. De Ramón Fernández A, Ruiz Fernández D, Sabuco García Y. Business process management for optimizing clinical processes: a systematic literature review. Health Informatics J 2020 Jun;26(2):1305-1320 [FREE Full text] [doi: 10.1177/1460458219877092] [Medline: 31581880]
14. openEHR Task Planning Specification. openEHR. URL: https://specifications.openehr.org/releases/PROC/latest/task_planning.html [accessed 2020-09-13]
15. Antibiotic guidelines. Johns Hopkins Medicine. URL: https://www.hopkinsmedicine.org/antimicrobial-stewardship/guidelines/ [accessed 2020-06-12]
16. Clinical Practice Guidelines Directions for a New Program. Washington, D.C., United States: National Academy Press; 1990.
17. Iglesias N, Juarez JM, Campos M. Comprehensive analysis of rule formalisms to represent clinical guidelines: selection criteria and case study on antibiotic clinical guidelines. Artif Intell Med 2020 Mar;103:101741. [doi: 10.1016/j.artmed.2019.101741] [Medline: 31928849]
18. Fox J, Johns N, Lyons C, Rahmanzadeh A, Thomson R, Wilson P. PROforma: a general technology for clinical decision support systems. Comput Method Program Biomed 1997 Sep;54(1-2):59-67. [doi: 10.1016/s0169-2607(97)00034-5]
19. Miksch S, Shahar Y, Johnson P. Asbru: a task-specific, intention-based, and time-oriented language for representing skeletal plans. In: Proceedings of the 7th Workshop on Knowledge Engineering: Methods & Languages (KEML-97). 1997 Presented at: 7th Workshop on Knowledge Engineering: Methods & Languages (KEML-97); 1997; Milton Keynes, UK.
20. Carbonell J, Etzioni O, Gil Y, Joseph R, Knoblock C, Minton S, et al. PRODIGY. SIGART Bull 1991 Jul;2(4):51-55. [doi: 10.1145/122344.122353]
21. Analysis of place/transition nets with timed arcs and its application to batch process control. In: Application and Theory of Petri Nets 1993. Berlin, Heidelberg: Springer; 1993.
22. Carl Adam P. Kommunikation mit Automaten. Universität Hamburg. URL: https://edoc.sub.uni-hamburg.de//informatik/volltexte/2011/160/ [accessed 2020-10-24]
23. Gatta R, Vallati M, Fernandez-Llatas C, Martinez-Millana A, Orini S, Sacchi L, et al. Clinical guidelines: a crossroad of many research areas. Challenges and opportunities in process mining for healthcare. In: Business Process Management Workshops. Cham: Springer; 2019.
24. Combi C, Oliboni B, Weske M, Zerbato F. Conceptual modeling of processes and data: connecting different perspectives. In: Conceptual Modeling. Cham: Springer; 2018.
25. Malek A, Raad I. Catheter- and device-related infections in critically ill cancer patients. In: Oncologic Critical Care. Cham: Springer; 2019.
26. Safdar N, Maki D, Mermel L. Infections due to infusion therapy. In: Bennett & Brachman's Hospital Infections, 5th Edition. Philadelphia, Pennsylvania, United States: Lippincott Williams and Wilkins; 2007.
27. Pascual A, Cercenado E, Salavert M, Elías García-Sánchez J, Eiros JM, Liñares J, et al. Update on pathogenesis and diagnosis of intravascular catheter-related infections. Enferm Infecc Microbiol Clin 2011 Mar;29 Suppl 4:16-21. [doi: 10.1016/S0213-005X(11)70032-5] [Medline: 21458716]
28. O'Grady NP, Alexander M, Burns LA, Dellinger EP, Garland J, Heard SO, Healthcare Infection Control Practices Advisory Committee (HICPAC). Guidelines for the prevention of intravascular catheter-related infections. Clin Infect Dis 2011 May;52(9):e162-e193 [FREE Full text] [doi: 10.1093/cid/cir257] [Medline: 21460264]
29. Tängdén T, Pulcini C, Aagaard H, Balasegaram M, Hara GL, Nathwani D, et al. Unavailability of old antibiotics threatens effective treatment for common bacterial infections. Lancet Infect Dis 2018 Mar;18(3):242-244. [doi: 10.1016/s1473-3099(18)30075-6]
30. Garnacho-Montero J, Aldabó-Pallás T, Palomar-Martínez M, Vallés J, Almirante B, Garcés R, et al. Risk factors and prognosis of catheter-related bloodstream infection in critically ill patients: a multicenter study. Intensive Care Med 2008 Dec;34(12):2185-2193. [doi: 10.1007/s00134-008-1204-7] [Medline: 18622596]

31. Manian F. IDSA guidelines for the diagnosis and management of intravascular catheter-related bloodstream infection. Clin Infect Dis 2009 Dec 01;49(11):1770-1; author reply 1771. [doi: 10.1086/648113] [Medline: 19891568]
32. Dijkman RM, Dumas M, Ouyang C. Semantics and analysis of business process models in BPMN. Inform Softw Technol 2008 Nov;50(12):1281-1294. [doi: 10.1016/j.infsof.2008.02.006]
33. Rodriguez-Loya S, Aziz A, Chatwin C. A service oriented approach for guidelines-based clinical decision support using BPMN. Stud Health Technol Inform 2014;205:43-47. [Medline: 25160142]
34. Combi C, Gozzi M, Posenato R, Pozzi G. Conceptual modeling of flexible temporal workflows. ACM Trans Auton Adapt Syst 2012 Jul;7(2):1-29. [doi: 10.1145/2240166.2240169]
35. Unified Modeling Language (UML). Object Management Group. URL: https://www.omg.org/spec/UML/2.5.1/PDF [accessed 2020-09-26]
36. Combi C, Oliboni B, Zerbato F. A modular approach to the specification and management of time duration constraints in BPMN. Inform Syst 2019 Sep;84:111-144. [doi: 10.1016/j.is.2019.04.010]
37. Cheikhrouhou S, Kallel S, Guermouche N, Jmaiel M. Towards a time-centric modeling of Business Processes in BPMN 2. In: Proceedings of the 15th International Conference on Information Integration andWeb-based Applications & Services (iiWAS'13). 2013 Presented at: 15th International Conference on Information Integration and Web-based Applications & Services (iiWAS'13); Dec, 2013; Austria URL: https://hal.archives-ouvertes.fr/hal-00921390 [doi: 10.1145/2539150.2539182]
38. Combi C, Oliboni B, Zerbato F. Modeling and handling duration constraints in BPMN 2.0. In: Proceedings of the Symposium on Applied Computing. 2017 Presented at: SAC 2017: Symposium on Applied Computing; Apr 3 - 7, 2017; Marrakech Morocco. [doi: 10.1145/3019612.3019618]
39. Combi C, Sala P, Zerbato F. Driving time-dependent paths in clinical BPMN processes. In: Proceedings of the Symposium on Applied Computing. 2017 Presented at: SAC 2017: Symposium on Applied Computing; Apr 3 - 7, 2017; Marrakech Morocco. [doi: 10.1145/3019612.3019620]
40. Expressiveness of temporal constraints for process models. In: Advances in Conceptual Modeling. Cham: Springer; 2018.
41. Gagne D, Trudel A. Time-BPMN. In: Proceedings of the 2009 IEEE Conference on Commerce and Enterprise Computing. 2009 Presented at: 2009 IEEE Conference on Commerce and Enterprise Computing; Jul 20-23, 2009; Vienna, Austria. [doi: 10.1109/cec.2009.71]
42. Watahiki K, Ishikawa F, Hiraishi K. Formal verification of business processes with temporal and resource constraints. In: Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics. 2011 Presented at: 2011 IEEE International Conference on Systems, Man, and Cybernetics; Oct 09-12, 2011; Anchorage, AK, USA. [doi: 10.1109/icsmc.2011.6083857]
43. Zerbato F. Tackling different business process perspectives: modeling data, time, and decisions in BPMN processes. University of Verona. 2019 Mar 26. URL: http://file:///C:/Users/user/Downloads/TESI_ZERBATO.pdf [accessed 2020-10-24]
44. Zerbato F, Oliboni B, Combi 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: 2015 International Conference on Healthcare Informatics; Oct 21-23, 2015; Dallas, TX, USA. [doi: 10.1109/ichi.2015.49]
45. Open industry specifications, models and software for e-health. openEHR Foundation. 2002. URL: https://www.openehr.org/ [accessed 2020-09-01]
46. openEHR task planning visual modelling language. openEHR Foundation. URL: https://specifications.openehr.org/releases/ [accessed 2020-12-13]
47. van der Aalst W, ter Hofstede A. YAWL: yet another workflow language. Inform Syst 2005 Jun;30(4):245-275. [doi: 10.1016/j.is.2004.02.002]
48. Case Management Model and Notation (CMMN). Object Management Group (OMG). URL: https://www.omg.org/cmmn/ [accessed 2020-09-26]
49. Decision Model and Notation (DMN). Object Management Group (OMG). URL: https://www.omg.org/dmn/ [accessed 2020-12-26]
50. Beale T. openEHR Decision Language Specification. openEHR Foundation. URL: https://specifications.openehr.org/releases/PROC/latest/decision_language.html [accessed 2020-09-13]
51. Beale T. openEHR Data Types Information Model. openEHR Foundation. URL: https://specifications.openehr.org/releases/RM/latest/data_types.html [accessed 2020-09-13]
52. Anani N. Exploring openEHR-based Clinical Guidelines in acute stroke care and research. Karolinska Institutet. URL: https://openarchive.ki.se/xmlui/handle/10616/44956 [accessed 2020-11-05]
53. Anani N, Mazya MV, Chen R, Prazeres Moreira T, Bill O, Ahmed N, et al. Applying openEHR's Guideline Definition Language to the SITS international stroke treatment registry: a European retrospective observational study. BMC Med Inform Decis Mak 2017 Jan 10;17(1):7 [FREE Full text] [doi: 10.1186/s12911-016-0401-5] [Medline: 28073358]
54. Zotero openEHR Group. openEHR Foundation. URL: https://www.zotero.org/groups/11839/openehr/items/RH8PH268/library [accessed 2021-02-23]

55. Alves DS, Maranhão PA, Pereira AM, Bacelar-Silva GM, Silva-Costa T, Beale TW, et al. Can openEHR represent the clinical concepts of an obstetric-Specific EHR - ObsCare Software? Stud Health Technol Inform 2019 Aug 21;264:773-777. [doi: 10.3233/SHTI190328] [Medline: 31438029]
56. Frexia F. Innovative information models to capture the dynamics of clinical processes: introducing process-oriented traceability in medical informatics specifications as a case study. Universit` a degli Studi di Cagliari. URL: http://hdl.handle.net/11584/260749 [accessed 2020-09-19]
57. Oliveira D, Miranda R, Abreu N, Leuschner P, Abelha A, Santos M, et al. Management of a pandemic based on an openEHR approach. Procedia Comput Sci 2020;177:522-527 [FREE Full text] [doi: 10.1016/j.procs.2020.10.072] [Medline: 35721473]
58. Braun R, Schlieter H. Requirements-based development of BPMN extensions: the case of clinical pathways. In: Proceedings of the 2014 IEEE 1st International Workshop on the Interrelations between Requirements Engineering and Business Process Management (REBPM). 2014 Presented at: 2014 IEEE 1st International Workshop on the Interrelations between Requirements Engineering and Business Process Management (REBPM); Aug 25, 2014; Karlskrona, Sweden. [doi: 10.1109/rebpm.2014.6890734]
59. Braun R, Burwitz M, Schlieter H, Benedict M. Clinical processes from various angles - amplifying BPMN for integrated hospital management. In: Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2015 Presented at: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Nov 09-12, 2015; Washington, DC, USA. [doi: 10.1109/bibm.2015.7359794]
60. Braun R, Esswein W. Classification of domain-specific BPMN extensions. In: The Practice of Enterprise Modeling. Berlin, Heidelberg: Springer; 2014.
61. Braun R, Schlieter H, Burwitz M, Esswein W. BPMN4CP: design and implementation of a BPMN extension for clinical pathways. In: Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2014 Presented at: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Nov 02-05, 2014; Belfast, UK. [doi: 10.1109/bibm.2014.6999261]
62. Braun R, Schlieter H, Burwitz M, Esswein W. BPMN4CP Revised -- extending BPMN for multi-perspective modeling of clinical pathways. In: Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS). 2016 Presented at: 2016 49th Hawaii International Conference on System Sciences (HICSS); Jan 05-08, 2016; Koloa, HI, USA. [doi: 10.1109/hicss.2016.407]
63. 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]
64. Beale T, Chen R. openEHR specification program. openEHR. URL: https://specifications.openehr.org/releases/PROC/latest/overview.html [accessed 2021-02-06]
65. TP-VML object library for draw. openEHR Foundation. URL: https://specifications.openehr.org/releases/PROC/latest/tp_vml/draw.io-lib/TP_VML_plans.xml [accessed 2021-02-05]
66. Extending BPMN 2.0: method and tool support. In: Business Process Model and Notation. Berlin, Heidelberg: Springer; 2011.
67. Allen JF. An interval-based representation of temporal knowledge. In: Proceedings of the 7th international joint conference on Artificial intelligence - Volume 1. 1981 Presented at: IJCAI'81: Proceedings of the 7th international joint conference on Artificial intelligence - Volume 1; Aug 24 - 28, 1981; Vancouver BC Canada.
68. openEHR Expression Language (EL). openEHR Foundation. URL: https://specifications.openehr.org/releases/BASE/Release1.0.4/expression.html [accessed 2020-12-13]
69. Wiemuth M, Junger D, Leitritz MA, Neumann J, Neumuth T, Burgert O. Application fields for the new object management group (OMG) standards case management model and notation (CMMN) and decision management notation (DMN) in the perioperative field. Int J Comput Assist Radiol Surg 2017 Aug 18;12(8):1439-1449. [doi: 10.1007/s11548-017-1608-3] [Medline: 28516301]

---

## Abbreviations

**BPM:** business process management
**BPMN:** business process model and notation
**CEN:** European Committee for Standardization
**CG:** clinical guideline
**CLT:** catheter lock therapy
**CMMN:** Case Management Model and Notation
**CP:** clinical pathway
**CR-BSI:** catheter-related bloodstream infection
**DL:** openEHR Decision Language
**DMN:** Decision Model and Notation
**EHR:** electronic health record
**GLIDE:** Graphic Language for Interactive Design

**HL7:** Health Level 7
**HRID:** human-readable identifier
**ISO:** International Organization for Standardization
**JHH:** Johns Hopkins Hospital
**OMG:** Object Management Group
**SHACL:** Shapes Constraint Language
**SPIN:** SPARQL Protocol and RDF Query Language Inferencing Notation
**SPARQL:** Simple Protocol and RDF Query Language
**ST:** systemic therapy
**TP:** openEHR Task Planning
**TP-VML:** Task Planning Visual Modeling Language
**YAWL:** Yet Another Workflow Language

*Edited by T Leung; submitted 26.04.21; peer-reviewed by I Gabashvili, Y Man; comments to author 21.07.21; revised version received 23.08.21; accepted 31.03.22; published 15.09.22.*

*Please cite as:*
*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;24(9):e29927*
*URL: https://www.jmir.org/2022/9/e29927*
*doi:10.2196/29927*
*PMID:36107480*

Original Paper

# The Clinical Effectiveness of Blended Cognitive Behavioral Therapy Compared With Face-to-Face Cognitive Behavioral Therapy for Adult Depression: Randomized Controlled Noninferiority Trial

Kim Mathiasen[1], PhD; Tonny E Andersen[2], PhD; Mia Beck Lichtenstein[1], PhD; Lars Holger Ehlers[3,4], PhD; Heleen Riper[5], PhD; Annet Kleiboer[5], PhD; Kirsten K Roessler[2], PhD

[1]Research Unit for Telepsychiatry and E-Mental Health, Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[2]Department of Psychology, University of Southern Denmark, Odense, Denmark
[3]Danish Center for Healthcare Improvements, Aalborg University, Aalborg, Denmark
[4]Nordic Institute of Health Economics, Aarhus, Denmark
[5]Department of Clinical, Neuro and Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

**Corresponding Author:**
Kim Mathiasen, PhD
Research Unit for Telepsychiatry and E-Mental Health
Department of Clinical Research
University of Southern Denmark
Heden 11
Odense, 5000
Denmark
Phone: 45 61677747
Email: kmathiasen@health.sdu.dk

## Abstract

**Background:** Internet-based cognitive behavioral therapy (iCBT) has been demonstrated to be cost- and clinically effective. There is a need, however, for increased therapist contact for some patient groups. Combining iCBT with traditional face-to-face (FtF) consultations in a blended format may produce a new treatment format (B-CBT) with multiple benefits from both traditional CBT and iCBT, such as individual adaptation, lower costs than traditional therapy, wide geographical and temporal availability, and possibly lower threshold to implementation.

**Objective:** The primary aim of this study is to compare directly the clinical effectiveness of B-CBT with FtF-CBT for adult major depressive disorder.

**Methods:** A 2-arm randomized controlled noninferiority trial compared B-CBT for adult depression with treatment as usual (TAU). The trial was researcher blinded (unblinded for participants and clinicians). B-CBT comprised 6 sessions of FtF-CBT alternated with 6-8 web-based CBT self-help modules. TAU comprised 12 sessions of FtF-CBT. All participants were aged 18 or older and met the diagnostic criteria for major depressive disorder and were recruited via a national iCBT clinic. The primary outcome was change in depression severity on the 9-item Patient Health Questionnaire (PHQ-9). Secondary analyses included client satisfaction (8-item Client Satisfaction Questionnaire [CSQ-8]), patient expectancy (Credibility and Expectancy Questionnaire [CEQ]), and working (Working Alliance Inventory [WAI] and Technical Alliance Inventory [TAI]). The primary outcome was analyzed by a mixed effects model including all available data from baseline, weekly measures, 3-, 6, and 12-month follow-up.

**Results:** A total of 76 individuals were randomized, with 38 allocated to each treatment group. Age ranged from 18 to 71 years (SD 13.96) with 56 (74%) females. Attrition rate was 20% (n=15), which was less in the FtF-CBT group (n=6, 16%) than in the B-CBT group (n=9, 24%). As many as 53 (70%) completed 9 or more sessions almost equally distributed between the groups (nFtF-CBT=27, 71%; nB-CBT=26, 68%). PHQ-9 reduced 11.38 points in the FtF-CBT group and 8.10 in the B-CBT group. At 6 months, the mean difference was a mere 0.17 points. The primary analyses confirmed large and significant within-group reductions in both groups (FtF-CBT: $\beta=-.03$; standard error [SE] 0.00; $P<.001$ and B-CBT: $\beta=-.02$; SE 0.00; $P<.001$). A small but significant interaction effect was observed between groups ($\beta=.01$; SE 0.00; $P=.03$). Employment status influenced the outcome differently between groups, where the B-CBT group was seen to profit more from not being full-time employed than the FtF group.

**Conclusions:** With large within-group effects in both treatment arms, the study demonstrated feasibility of B-CBT in Denmark. At 6 months' follow-up, there appeared to be no difference between the 2 treatment formats, with a small but nonsignificant difference at 12 months. The study seems to demonstrate that B-CBT is capable of producing treatment effects that are close to FtF-CBT and that completion rates and satisfaction rates were comparable between groups. However, the study was limited by small sample size and should be interpreted with caution.

**Trial Registration:** ClinicalTrials.gov NCT02796573; https://clinicaltrials.gov/ct2/show/NCT02796573

**International Registered Report Identifier (IRRID):** RR2-10.1186/s12888-016-1140-y

***(J Med Internet Res 2022;24(9):e36577)*** doi:10.2196/36577

**KEYWORDS**

depression; depressive disorder, major; cognitive therapy; CBT; treatment outcome; blended care; blended cognitive behavioral therapy; effectiveness; Denmark

## Introduction

### Background

Depression is a prevalent and disabling disorder with a high risk of relapse and large individual and societal costs [1-6]. Effective treatments do exist [7], although a large gap is seen between the need for and use of treatments [8]. This gap has led researchers to explore alternative modes of treatment delivery. One such novel treatment format is internet-based cognitive behavioral therapy (iCBT) [9-17], in which the patient is administered access to an online treatment program based on CBT. The highest clinical effect is seen when clinical guidance is provided during the course of treatment [18-20]. However, despite the evidence for the effect of guided iCBT, there is a need for increased therapist contact among some patient groups as well as a need to provide a treatment format, which is more compatible with, and thus easier to implement in, the existing health care services [21-24].

Combining iCBT with traditional face-to-face (FtF) consultations in a blended CBT format (B-CBT), in which both online components and FtF sessions are included in 1 coherent CBT protocol, may alleviate some of the difficulties associated with iCBT for depression, while preserving some of the advantages of both iCBT and FtF-CBT alike. First, by including FtF sessions, the therapist can individualize the therapy taking the idiosyncratic case formulation of the patient, the specific disorder, and possible comorbidity into account. Second, as B-CBT in the format tested in this study only provides half the number of sessions as traditional FtF-CBT, the capacity of the treating clinician is increased compared with traditional CBT. Third, the burden and cost of travel by the patient can be reduced compared with FtF-CBT. Fourth, the online modules are available at the time and place needed by the patients—and they can be re-viewed multiple times. Fifth, the inherently structured format of the online modules ensures high treatment fidelity, for example, by delivering the same psychoeducation and exercises to all patients. Sixth, one of the principal barriers for the uptake of iCBT seems to be skepticism concerning allotting the majority of therapy to a computer [25], a barrier possibly alleviated by the B-CBT [26]. Finally, the blended format is more compatible with the existing health care services and as a consequence should be easier to implement than iCBT [27].

Few studies have investigated the use of blended care combining internet-based psychotherapeutic modules and FtF sessions into 1 coherent treatment manual to treat adult depression [28-31]. Generally, however, they do indicate positive outcomes. In a randomized controlled trial conducted in primary care in Tromsø, Norway, clinical psychologists delivered 30-minute sessions following each online module [29]. They were able to document a significant difference with a moderate to large effect size ($d$=0.65) on depressive symptoms (Beck Depression Inventory II [BDI-II]) favoring blended care over waiting list. The intervention predominantly received positive evaluations suggesting acceptability and satisfaction with the treatment. In addition, a qualitative study found that the FtF consultations increased motivation to persist with the iCBT program [32]. Another recent example is the development and initial evaluation of a program for B-CBT in The Netherlands. This was tested at an outpatient clinic of a specialized mental health care center in Amsterdam. The study was designed as a feasibility study and included only 9 patients. However, the patients perceived the intervention as positive, although the authors rightly noted that no conclusion can be derived from such a small sample [31]. A cohort study from the United States found a significant and large reduction in symptoms of depression but had no comparison group [33].

In this study we compared directly the clinical effect on adult depression of B-CBT and FtF-CBT in a randomized, controlled, noninferiority study in parallel groups, recruiting from a routine care iCBT clinic in the Region of Southern Denmark.

### Aims and Hypotheses

The primary aim of this study was to compare the clinical effectiveness of B-CBT for major depressive disorder in adults with treatment as usual (TAU) defined as 12 sessions of FtF-CBT. It is hypothesized that B-CBT will be no less clinically effective than FtF-CBT, and that it will be acceptable and satisfactory to patients and clinicians.

# Methods

## Design

The study was a randomized, controlled, noninferiority trial comparing B-CBT with FtF-CBT. It was part of the research program e-Mental Health Research (ENTER) located in and coordinated from the Centre for Telepsychiatry in the Mental Health Services of Southern Denmark, Odense. Additionally, this study was affiliated with the European Union (EU) study E-COMPARED [34]. However, the E-COMPARED study ended prior to this, and thus does not include the total sample. In this article we explore the full data set of the trial.

## Ethics Approval

The trial was approved by the Ethics Committee of the Region of Southern Denmark (registration number S-20150150) prior to instigation. The trial followed the Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects [35]. All participants received both written and oral information about the trial and signed written informed consent before entering the study.

## Trial Registration

The trial was registered with ClinicalTrials.gov NCT02796573. The trial protocol was published previously [36].

## Study Funding

Funding was granted from the Research Fund of the Mental Health Services of Southern Denmark, and from the Innovation Fund Denmark, as part of the project ENTER (ID: 5159-00002B). Both are public funds. None of the funds have had any role in the design of the study nor in the collection, analysis or interpretation of the data, or writing of the manuscript.

## Participants

### *Eligibility Criteria*

All participants were 18 years of age or older and met the diagnostic criteria for major depressive disorder according to the Diagnostic and Statistical Manual of Mental Disorders 4th edition text revision (DSM-IV-TR) [37] as assessed by clinical psychologists. The diagnosis was confirmed by the research team using the semistructured interview Mini-International Neuropsychiatric Interview version 5.0 (M.I.N.I.) [38]. Furthermore, a score of at least five on the 9-item Patient Health Questionnaire-9 (PHQ-9) [39,40] was required. Patients were excluded in case of current high risk of suicide or if they had a comorbid substance dependence, bipolar disorder, psychotic illness, or obsessive-compulsive disorder. Additionally, participants were excluded if they concurrently received psychological treatment for depression. They were also required to comprehend the Danish language and have access to a personal computer and internet connection. Finally, they needed to be able and willing to travel to the physical location of the trial even if they were randomized to the FtF condition.

### *Recruitment*

Participants were recruited from March 1, 2016, to April 1, 2018, from the iCBT clinic “Internetpsykiatrien,” which is situated within secondary mental health care (Centre for Telepsychiatry) at the Mental Health Services of the Region of Southern Denmark [41,42]. Internetpsykiatrien offers guided iCBT treatment for anxiety and depression with self-referral. Psychologists or master students in psychology interviewed all participants using M.I.N.I. [38] to confirm diagnosis. The interviews were administered either FtF or by telephone. In case the participants were on antidepressant medication, they were asked to keep it stable during treatment if possible. They were asked to report any changes in medication to the research team. Access to the program was provided by the research team.

## Randomization and Blinding

An independent researcher from the EU study E-COMPARED [34], who was not involved in the trial, performed the randomization at an individual level, stratified by country after eligibility and baseline measurement. A random number generator (Random Allocation Software) was applied with an allocation ratio of 1:1. Block randomization was used with block sizes varying from 8 to 14 allocations per block.

It was not possible to blind the patients nor the treating clinicians to the allocated treatment. However, those assessing the participants were blinded to allocation as were the researchers and statisticians involved up until the point of interpretation of the results. Some questionnaires were only administered to the B-CBT group and were kept in a separate data set.

## Interventions

### *Blended Treatment (B-CBT)*

In the blended condition, 6 individual FtF-CBT sessions were alternated with 6-8 online CBT modules delivered through an internet-based treatment program. The FtF consultations were provided by a psychologist at the Centre for Telepsychiatry with physical presence by the participants and the therapists.

The program (NoDep) was previously developed (2015) as part of a public private innovation project between The Region of Southern Denmark and Context Consulting. It was based on CBT for depression and included 6 mandatory modules and 2 optional ones. The core components of the mandatory modules were psychoeducation, cognitive restructuring, behavioral activation, behavior experiments, and relapse prevention. The optional modules comprised coping with rumination and restructuring of core beliefs. All online modules were introduced in the FtF sessions. Modules the participants had previously worked with could be addressed in the FtF sessions if needed. The decision as to whether any optional modules need to be added was taken jointly by the patient and the psychologist based on patient needs, motivation, and possible time constraints. See Table 1 for an overview of the intervention. All modules were delivered via multimedia elements including video, audio, interactive exercises, calendar, and PDF summaries. The program had a build-in workflow predetermining the order in which the modules were presented. All data were stored in Europe and encrypted during storage and transmission.

**Table 1.** Overview of interventions.

| Intervention and session number | Format of delivery | Content | Example of exercise |
|---|---|---|---|
| **B-CBT**[a] | | | |
| 1 | FtF[b] | Introduction and psychoeducation about depression and the treatment | Find a helper |
| 2 | Online module | Introduction to the program, psychoeducation about depression, and goals for the treatment | Problem/goal list |
| 3 | FtF | Idiosyncratic model of the disorder | Cognitive case formulation |
| 4 | Online module | Psychoeducation about behavior in depression | Activity registration |
| 5 | FtF | Accordance between personal values and behavior. Introduction to cognitive restructuring | Simple exercise for cognitive restructuring |
| 6 | Online module | Changing behavior based on activity registration and personal values | Activity planning |
| 7 | FtF | Psychoeducation about negative automatic thoughts and cognitive restructuring | Cognitive restructuring exercise |
| 8 | Online module | Psychoeducation about negative automatic thoughts and cognitive restructuring | Cognitive restructuring exercise |
| 9 | FtF | Psychoeducation about behavioral experiments. Decision is made as to whether to include either or both of the extra modules | Behavioral experiment |
| 10 | Online module (A, B) | Behavioral experiments (A: psychoeducation about core beliefs, B: coping with rumination) | Behavioral experiment (A: challenge core beliefs; B: test 3 techniques for coping with rumination) |
| 11 | FtF | Summing up, relapse prevention | Continuation of preferred exercises |
| 12 | Online module | Summing up, relapse prevention | Personal relapse prevention plan |
| **TAU**[c] | | | |
| 1 | FtF | Introduction and psychoeducation about depression and the treatment | Find a helper |
| 2 | FtF | Psychoeducation and goals for the treatment | Problem/goal list |
| 3 | FtF | Idiosyncratic model of the disorder | Cognitive case formulation |
| 4 | FtF | Psychoeducation about behavior in depression | Activity registration |
| 5 | FtF | Accordance between personal values and behavior. Introduction to cognitive restructuring | Simple exercise for cognitive restructuring |
| 6 | FtF | Changing behavior based on activity registration and personal values | Activity planning |
| 7 | FtF | Psychoeducation about negative automatic thoughts and cognitive restructuring | Cognitive restructuring exercise |
| 8 | FtF | Psychoeducation about negative automatic thoughts and cognitive restructuring | Cognitive restructuring exercise |
| 9 | FtF | Psychoeducation about behavioral experiments | Behavioral experiment |
| 10 | FtF | Psychoeducation about core beliefs or continue working on behavioral experiments | Challenge core beliefs or behavioral experiment |
| 11 | FtF | Psychoeducation about rumination or beginning of relapse prevention | Test 3 techniques to cope with rumination or start personal relapse prevention plan and continuation of preferred exercise |
| 12 | FtF | Summing up, relapse prevention | Personal relapse prevention plan |

[a] B-CBT: blended cognitive behavioral therapy.

[b] FtF: face-to-face.

[c] TAU: treatment as usual.

To provide technical support to the participants, the existing procedures at the Centre for Telepsychiatry were used, which consisted of 2 levels: the first was handled by the clinicians, the second went through an error report system to the company that provided the software (Context Consulting).

No important changes were made to the program or the protocol during the trial.

### *Treatment as Usual*

TAU defined as 12 sessions of FtF-CBT was also provided by a psychologist at the Centre for Telepsychiatry with physical presence and comprised the same core components as the B-CBT condition. Additionally, interventions on core beliefs and rumination could be included according to the same criteria as in the B-CBT condition. See Table 1 for an overview of the intervention.

Both treatment conditions were described in a single common treatment protocol, thus ensuring similar treatment content and order of interventions across the 2 groups. They were both intended to last approximately 12 weeks.

### *Safety Procedures*

Patients in either condition were monitored weekly for symptoms of depression including suicidal ideation and intent. In case a participant's condition deteriorated or showed signs of suicidal intent, a standard assessment procedure used in all of the secondary mental health care services in the Region of Southern Denmark was conducted. The patient was discontinued if necessary and referred to other relevant treatment.

### *Adherence and Fidelity*

Licensed clinical psychologists or psychologists under supervision of the primary researcher (KM), who is also a licensed clinical psychologist, delivered all FtF consultations. To assess clinician fidelity [43], all FtF sessions were audio recorded and 20 sessions were randomly selected and evaluated by an external clinical expert (clinical psychologist and PhD with many years' experience). Clinician adherence was defined as the number of *pre*scribed interventions that were *pro*scribed in the session. The level of agreement between the 2 were rated on a 5-point scale ranging from none (1) to all (5) [44].

To increase adherence, participants received automated reminders of homework assignments and questionnaires. Furthermore, in case a participant was inactive, he or she would be contacted by telephone or email. Additionally, in case a participant was unwilling or unable to engage with the program at home, a computer was set up at the clinic, for participants to engage with the online program on-site. This was never used, however.

## Outcome Measurements

After consent was granted, baseline measures were administered prior to randomization. Follow-up measurements were conducted 3, 6, and 12 months after baseline. Additionally, weekly measures were provided during treatment. The questionnaire packages were administered online using a secure web application for building and managing online surveys (REDCap), except for the weekly monitoring of the B-CBT group, for which the packages were administered automatically by the treatment program.

Data were stored by the Odense Patient data Exploratory Network (OPEN) [45]. Data were collected, transferred, and stored securely electronically as approved by the Danish Data Protection Agency (journal number: 14/26634, registration number: 2008-58-0035).

The PHQ-9 [39] was used as the primary outcome measure. The PHQ-9 is a 9-item questionnaire developed to measure depressive symptomatology in the primary health care sector. The 9 items are each scored on a 0-3-point scale with the total score ranging from 0 to 27, with higher scores indicating more severe depression. The authors suggest using cut-off points of 5, 10, and 15 for mild, moderate, and severe levels of depression, respectively, in the guide to the instrument substantiated by a review [40]. The PHQ-9 has been shown to have good psychometric properties [46].

A number of additional measures were administered to assess different aspects of the participants' symptomatology and experience during the treatments. The 16-item Quick Inventory of Depressive Symptomatology Self-Report (QIDS-16-SR) [47,48] was used in addition to the PHQ-9 because it is a promising questionnaire for assessing depressive symptoms, especially in specialized mental health care and to conduct secondary analyses of primary latent construct of interest: depression. To measure the participants' satisfaction with the treatments, the 8-item Client Satisfaction Questionnaire (CSQ-8) [49,50] was used. The Credibility and Expectancy Questionnaire (CEQ) [51] was used to measure the participants' expectancy and judgment of credibility of the treatments. Finally, the level of therapeutic alliance was measured using the Working Alliance Inventory-Short Revised (WAI-SR) [52-54] and was rated by both the participants and the clinicians. For further description of the measures used, we refer to Mathiasen et al [36].

## Statistical Analyses

### *Baseline Characteristics*

Characteristics of the sample at baseline was described using descriptive statistics and compared across groups using unpaired *t* tests for continuous variables and chi-square tests for categorical variables. If continuous variables violated the assumption of normality, nonparametric tests were used (Kruskal-Wallis/Wilcoxon signed-rank test). In cases of small cell sizes, exact tests were used (Fisher exact test).

### *Primary Analysis*

For the primary analyses a linear multilevel mixed effects model with restricted maximum likelihood estimator was used as intention-to-treat analyses. PHQ-9 scores were used as response variable. Time was included as a fixed effect and as a random effect nested within participant (random slope and intercept) [55]. Correlation between slope and intercept was assumed. All available data were included. Missing values were handled by use of mixed effects models including all available data.

All inferences assumed normally distributed error terms and heteroscedasticity, which were substantiated by visual inspection of a q-q normality plot and a plot of fitted values versus standardized residuals.

Remission was defined as a score of <5 on the PHQ-9. Response to treatment was defined as 50% or more reduction on the PHQ-9.

The noninferiority margin was set to *d*=0.2.

### *Acceptability*

Acceptability was estimated from measures of client satisfaction (CSQ-8) and working alliance as reported by the participants (WAI-SR and Technical Alliance Inventory [TAI]) and the clinicians (WAIc). Means were compared across groups using unpaired *t* tests on raw scores using case-wise deletion in case of missing data.

### *Predictor Analyses*

Mixed effects models using all available data were applied for analyses of interactions between group and baseline variables by the intention-to-treat principle. One model per predictor was used with PHQ-9 as the response variable in a series of univariate analyses. This was done to test whether baseline characteristics affected outcome differently in the 2 treatments. Inclusion of all parameters would have overfitted the model due to sample size. Time was included as both a fixed effect and a random effect nested in individuals (similar to the primary analysis).

Second, analyses of predictors of symptomatic change in the total sample were also conducted using a mixed effects model with PHQ-9 as response variable. Both multivariate and a series of univariate analyses were conducted. No group interaction was included in these analyses.

### *Completion*

Having completed 9 or more (75%) sessions (out of 12) was counted as completion and mean completion rates were compared between groups by unpaired *t* test. The completion rate of the B-CBT group included the sum of online modules and FtF sessions attended.

To assess the odds of noncompletion predicted from the participants' baseline characteristics, a multivariate logistic regression analysis was conducted. As the response variable, a dichotomous variable for completion was used. Additionally, univariate logistic regression analyses were conducted using 1 model per predictor to investigate whether noncompletion was predicted differently between the FtF-CBT treatment and the B-CBT, which included an interaction term with group.

All calculations were performed using R version 3.4.4 (R Foundation for Statistical Computing) [56]. Mixed effects linear models were calculated using the ImerTest package [57], which fits models by use of the lme4 package [58] and provides *P* values by use of the Satterthwaite degrees of freedom method. Two-way analyses were used with *P*<.05 as the threshold for significance for inferential statistics. All CIs were calculated by bootstrapping using boot.ci [57].

## Results

### Description of Participants

Table 2 shows the baseline characteristics of all participants and Figure 1 shows the patient flow. In total, 76 were randomized, with 38 allocated to each group. Attrition was somewhat unevenly distributed between groups with 8 being lost to follow-up in the FtF-CBT group and 16 in the B-CBT group. Nonetheless, due to the weekly measurement scheme and the use of mixed effects models, all but one was included in the primary analyses.

The included sample was predominantly female (56/76, 74%) and young with a mean age of 35.0 (SD 13.96) years (median 30 years), although a large age range was seen (18-71 years). Most had moderate to highly severe levels of depression (66/76, 87%) with a mean score of 15.25 (SD 4.04) on the PHQ-9.

No significant differences were observed between the 2 groups on baseline characteristics except for scores on the CEQ measuring the participants' expectations and credibility of the treatments (see Table 2 for *P* values). The participants in the B-CBT group scored lower on treatment credibility and expectancy of treatment outcome. This raised suspicion as to whether the difference could have been caused by the participants being aware of their group allocation prior to responding to the questionnaire. However, when investigated, it did not seem to be the case. Likewise, no obvious outliers were driving the difference and the distribution of scores seemed reasonable upon visual inspection. A sensitivity analysis of the primary analysis was conducted controlling for the credibility and expectancy scores, but it did not change the outcome.

Among the included sample, 7 were on the brink of violating exclusion criteria, 3 were in psychological treatment at the point of assessment, 2 had some obsessive compulsive disorder symptoms, and 2 were not depressed according to MINI, but scored 9 and 17 on the PHQ-9, respectively. When comparing analyses including or excluding these cases, the outcome did not change. To avoid causing any changes to the analysis plan, all analyses were performed including these participants.

**Table 2.** Characteristics of participants (N=76)[a].

| Characteristics | FtF-CBT[b] | B-CBT[c] | *P* value |
|---|---|---|---|
| **Baseline description** | | | |
| Age, mean (SD) | 35.16 (14.14) | 34.78 (13.98) | .91 |
| Female gender, n/N (%) | 29/37 (78) | 27/37 (73) | .79 |
| PHQ-9[d], mean (SD) | 16.05 (3.83) | 14.42 (4.14) | .08 |
| Credibility, mean (SD) | 0.67 (2.01) | –0.69 (2.28) | .009[e] |
| Expectancy, mean (SD) | 0.70 (2.22) | –0.72 (2.88) | .02[f] |
| **Marital status, n/N (%)** | | | |
| Single | 13/37 (35) | 14/37 (38) | |
| Divorced | 5/37 (14) | 6/37 (16) | |
| Widow/widower | 0/37 (0) | 0/37 (0) | |
| Cohabiting | 9/37 (24) | 8/37 (22) | |
| Married | 10/37 (27) | 8/37 (22) | |
| Prefer not to answer | 0/37 (0.0) | 1/37 (3) | |
| **Highest education, n/N (%)** | | | |
| Further education <3 years | 7/37 (19) | 8/37 (22) | |
| Further education 3-4 years | 13/37 (35) | 13/37 (35) | |
| Higher education >4 years | 4/37 (11) | 3/37 (8) | |
| Fundamental school <8 years | 0/37 (0) | 0/37 (0) | |
| Fundamental school 9-10 years | 3/37 (8) | 3/37 (8) | |
| Gymnasium (3 years) | 9/37 (24) | 5/37 (14) | |
| Skilled worker | 1/37 (3) | 5/37 (14) | |
| **Employment status, n/N (%)** | | | .34 |
| Full-time employed | 9/36 (25) | 4/34 (12) | |
| Part-time employed | 5/36 (14) | 9/34 (27) | |
| Sick leave | 11/36 (31) | 9/34 (27) | |
| Leave of absence | 2/36 (6) | 0/34 (0) | |
| Retired | 1/36 (3) | 1/34 (3) | |
| Unemployed | 8/36 (22) | 11/34 (32) | |
| **Treatment preference, n/N (%)** | | | .82 |
| No preference | 16/37 (43) | 18/36 (50) | |
| Blended care | 9/37 (24) | 7/36 (19) | |
| Face-to-face | 12/37 (32) | 11/36 (31) | |
| **Depression severity, n/N (%)** | | | |
| No | 0/37 (0) | 0/36 (0) | |
| Mild | 3/37 (8) | 4/36 (11) | |
| Moderate | 9/37 (24) | 14/36 (39) | |
| Severe | 19/37 (51) | 16/36 (44) | |
| Highly severe | 6/37 (16) | 2/36 (6) | |

[a] Percentages calculated considering attrition.

[b] FtF-CBT: face-to-face cognitive behavioral therapy.

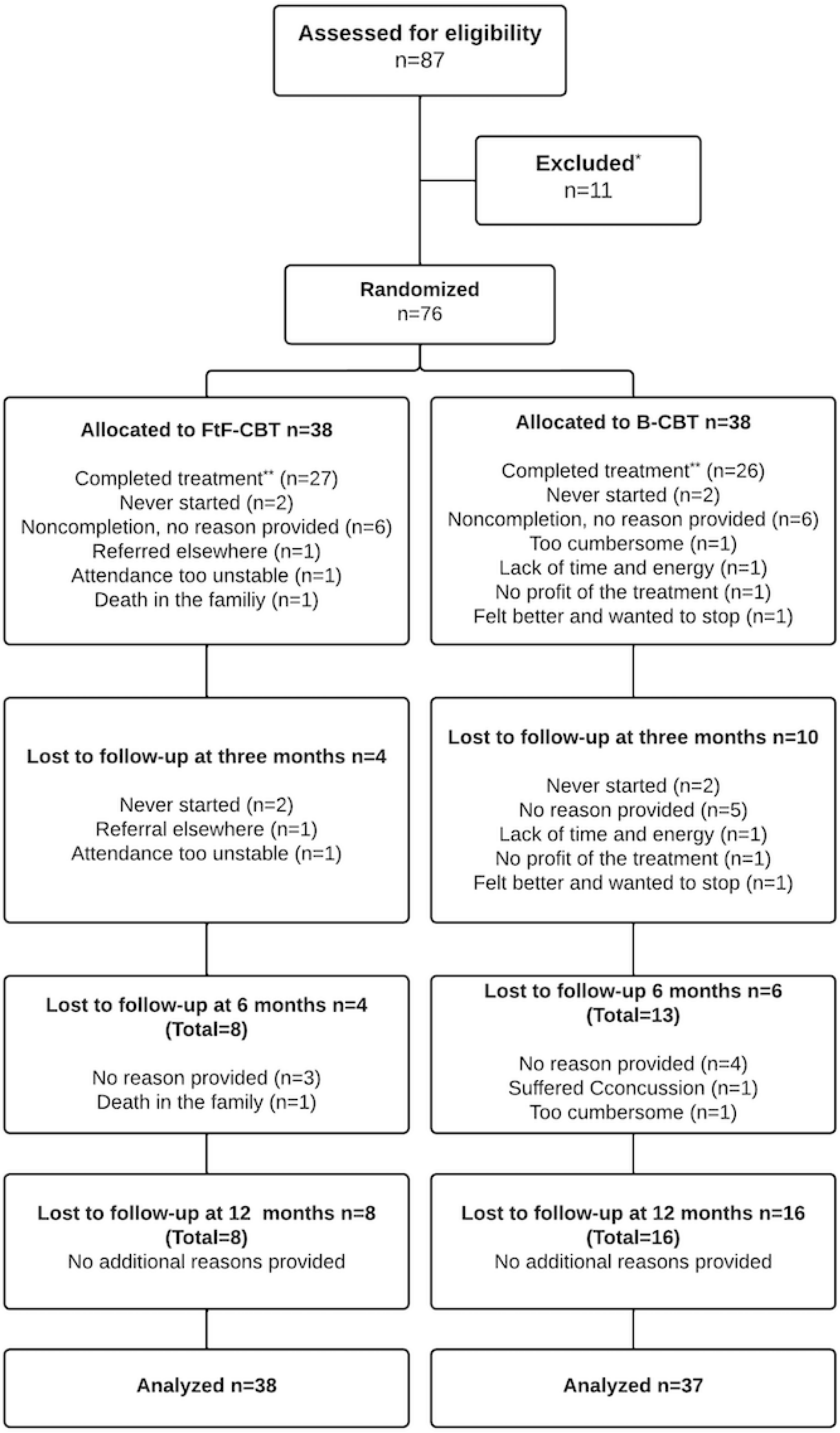
[c] B-CBT: blended cognitive behavioral therapy.

[d]PHQ-9: 9-item Patient Health Questionnaire.

[e]P<.01.

[f]P<.05.

**Figure 1.** Patient flow. *We did not store any data on any patients who had not provided informed consent. Consequently, no reasons can be provided for this category. **Treatment was regarded as completed when more than 9 sessions were completed. B-CBT: blended cognitive behavioral therapy; FtF-CBT: face-to-face cognitive behavioral therapy.

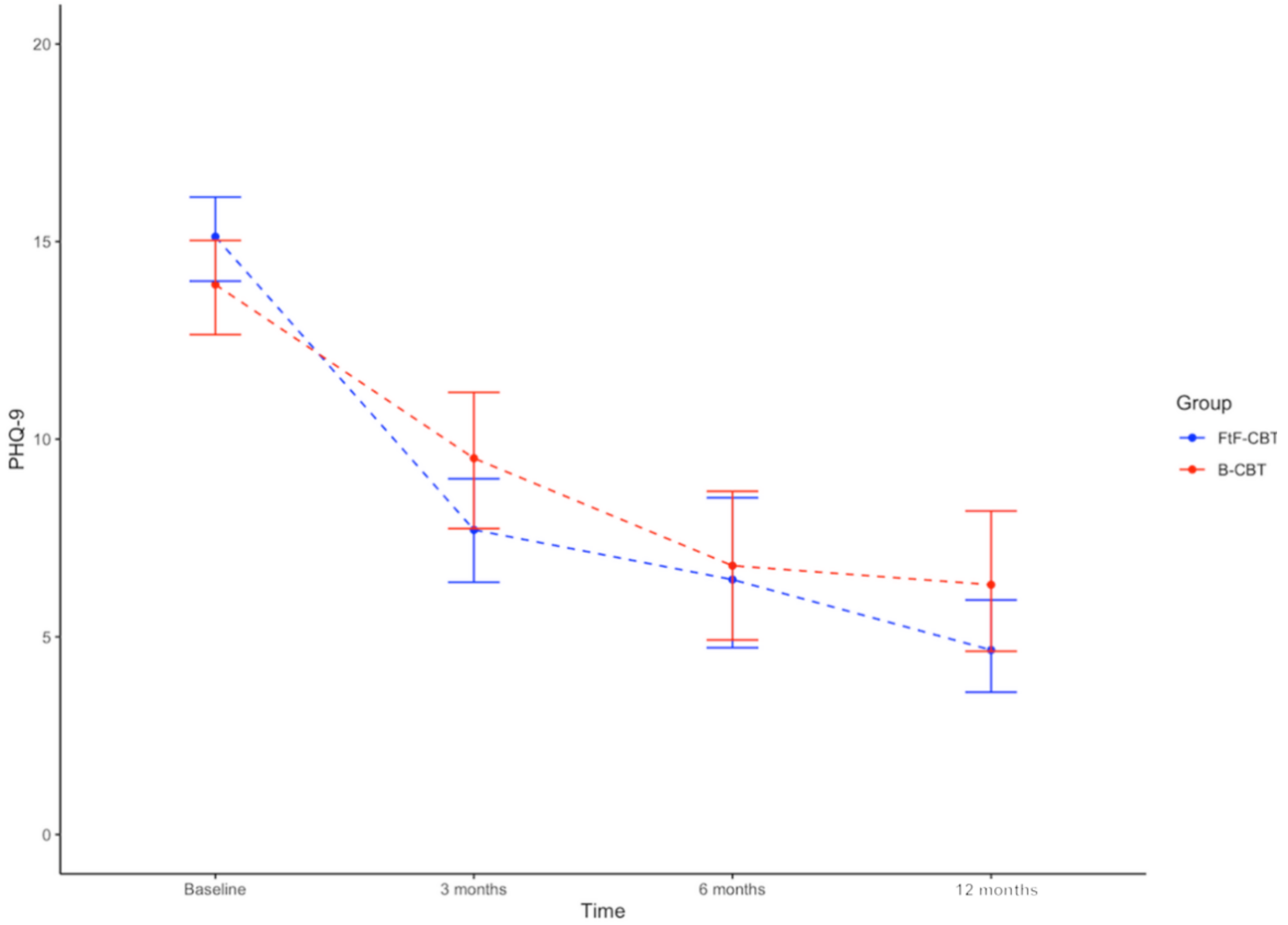


## Treatment Effect

### Overview

Initially, we report observed means (Figure 2), standard errors (SEs), and standardized mean differences (Cohen *d*) on the primary outcome measure (PHQ-9). Following this we report results of the mixed effects models.

**Figure 2.** Change in depression on PHQ-9. B-CBT: blended cognitive behavioral therapy; FtF-CBT: face-to-face cognitive behavioral therapy; PHQ-9: 9-item Patient Health Questionnaire.



### Observed Means

In both groups, large changes in the mean scores within groups were observed on the primary outcome measure (PHQ-9; Table 3). Within the FtF-CBT group, the mean score decreased from 16.05 (SE 0.63) at baseline to 4.67 (SE 0.62) at 12 months' follow-up. Likewise, in the B-CBT group, the mean score reduced from 14.42 (SE 0.69) to 6.32 (SE 0.95). In both groups the within-group changes in mean scores from baseline to 12-month follow-up revealed large, standardized effect sizes ($d_{\text{FtF-CBT}}$=–2.04, $d_{\text{B-CBT}}$=–1.57) [58-61].

Between groups, a trend in effect size was noted favoring the FtF-CBT group at 3 months' follow-up (*d*=–0.5, CI –1.62 to 0.62) but not at 6 months (*d*=0.03, CI –1.43 to 1.49), where the difference had all but disappeared, amounting to just 0.17 points on the PHQ-9 and stayed well within the noninferiority margin of *d*=0.2. At 12 months' follow-up, a difference could be observed slightly favoring FtF-CBT (*d*=–.42, CI –1.49 to 0.65). However, at all measurement points, the CIs were overlapping and were stretching beyond the noninferiority margin, rendering it impossible to infer generalizability of the results of noninferiority. A similar picture was seen on the secondary outcome of the QIDS (Figure 3).

**Table 3.** Observed means for PHQ-9[a].

| Timepoint | FtF-CBT[b] |  | B-CBT[c] |  |
|---|---|---|---|---|
|  | Mean (SE[d]) | *d* (CI) | Mean (SE) | *d* (CI) |
| **Baseline** |  |  |  |  |
| Mean | 16.05 (0.63) |  | 14.42 (0.69) |  |
| **Three months** |  |  |  |  |
| Mean | 7.71 (0.7) |  | 9.93 (0.92) |  |
| Between-groups effect size[e] |  | –0.5 (–1.62 to –1.17) |  |  |
| Within-group effect size[f] |  | –2.04 (–2.91 to –1.17) |  | –1.57 (–2.68 to –0.46) |
| **Six months** |  |  |  |  |
| Mean | 6.97 (1.09) |  | 6.8 (0.99) |  |
| Between-groups effect size[e] |  | 0.03 (–1.43 to 1.49) |  |  |
| Within-group effect size[f] |  | –2.09 (–3.29 to –0.89) |  | –1.52 (–2.67 to –0.37) |
| **Twelve months** |  |  |  |  |
| Mean | 4.67 (0.62) |  | 6.32 (0.95) |  |
| Between-groups effect size[e] |  | –0.42 (–1.49 to 0.65) |  |  |
| Within-group effect size[f] |  | –2.04 (-2.91 to –1.17) |  | –1.57 (–2.68 to –0.46) |

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

[b] FtF-CBT: face-to-face cognitive behavioral therapy.

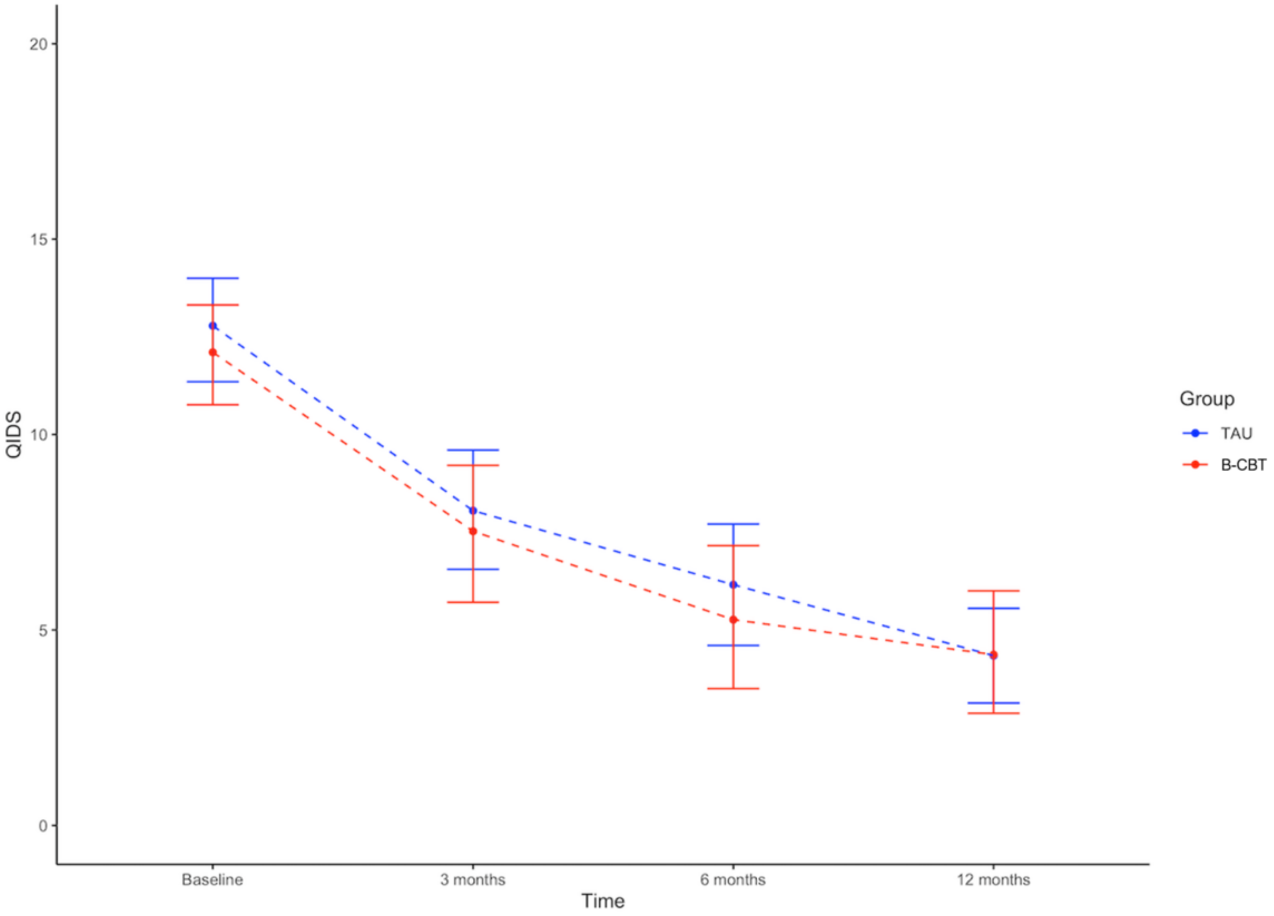
[c] B-CBT: blended cognitive behavioral therapy.

[d] SE: standard error.

[e] Independent samples [60].

[f] Formula 3 in Dunlap et al [61] for dependent samples.

**Figure 3.** Change in depression on QIDS. B-CBT: blended cognitive behavioral therapy; QIDS: Quick Inventory of Depressive Symptomatology; TAU: treatment as usual.



### Primary Analyses

As can be seen in Table 4, the primary analyses using linear mixed effects models with the PHQ-9 as outcome variable confirmed the within-group improvements in both groups being significant (FtF-CBT: β=–.03; SE 0.00; *P*<.001 and B-CBT: β=–.02; SE 0.00; *P*<.001), which was also the case for the QIDS scores (FtF-CBT: β=–.02; SE 0.00; *P*<.001 and B-CBT: β=–.01; SE 0.00; *P*<.001). The β values are small, as they represent the change in the outcome measure per day. Between groups, a very small but significant interaction effect was observed on the PHQ-9 (β=.01; SE 0.00; *P*=.03), indicating a slight advantage of the FtF-CBT group. However, this was not the case on the QIDS (β=.01; SE 0.00; *P*=.05), which was just above the significance level.

A negative correlation was observed between intercept and slope in the primary model (*r*=–0.29), indicating that a higher initial score (intercept) correlated with a steeper negative slope (symptomatic improvement).

**Table 4.** Results of the mixed effects linear regressions.

| Fixed effects | Primary analysis | | | Within-group FtF-CBT[a] | | | Within-group B-CBT[b] | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimates | SE[c] | *P* value | Estimates | SE | *P* value | Estimates | SE | *P* value |
| (Intercept) | 12.71 | 0.61 | $<2 \times 10^{-16}$[d] | 12.71 | 0.60 | $2 \times 10^{-16}$[d] | 12.47 | 0.64 | $<2 \times 10^{-16}$[d] |
| Time | –0.03 | 0.00 | $3.12 \times 10^{-16}$[d] | –0.03 | 0.00 | $2.24 \times 10^{-11}$[d] | –0.02 | 0.00 | $3.59 \times 10^{-9}$[d] |
| Group B-CBT | –0.23 | 0.89 | .80 | N/A[e] | N/A | N/A | N/A | N/A | N/A |
| Time × group B-CBT | 0.01 | 0.00 | .03[f] | N/A | N/A | N/A | N/A | N/A | N/A |

[a] FtF-CBT: face-to-face cognitive behavioral therapy.

[b] B-CBT: blended cognitive behavioral therapy.

[c] SE: standard error.

[d] *P*<.001.

[e] N/A: not applicable.

[f] *P*<.05.

### Acceptability

There was no significant difference in client satisfaction between groups (mean difference –2.18; $t_{39.36}$=2.16; *P*=.15). Furthermore, no significant difference in working alliance was observed when reported by the participants (mean difference 2.31; $t_{50.08}$=1.14; *P*=.26). However, the difference between groups was larger and significant when rated by the treating clinicians (mean difference 6.27; $t_{58.51}$=3.68; *P*<.001).

The rates of participants responding to treatment at 12 months based on the PHQ-9 were 83% (25/30) in the FtF-CBT group and 64% (14/22) in the B-CBT group. The remission rates at 12 months were 60% (18/30) for the FtF-CBT group and 50% (11/22) for the B-CBT group. When inspecting all individual slopes of the primary model, we found no negative individual slopes, indicating that none of the participants' depressive condition deteriorated.

Finally, 20 randomly selected audio-recorded sessions were examined for treatment fidelity by an external expert in clinical psychology. Among the sample, session numbers ranged from 3 to 12, 3 of 4 therapists were represented, and both groups were well represented, with 14 sessions being from the FtF-CBT group. The mean score of treatment fidelity was 4.25 (SD 0.71) on a scale ranging from 1 (not compliant with the protocol) to 5 (completely compliant with the protocol).

### Predictor Analyses

In a multivariate analysis of the total sample, only being on sick leave and preferring blended care predicted outcome. Being on sick leave added to the slope (3.96; SE 1.54; *P*=.02), that is, produced a smaller reduction in symptom change. Preferring blended care subtracted from the slope estimate (–3.25; SE 1.53; *P*=.04), thus signifying an increase in symptom reduction. Table 5 summarizes all predictor variables with SEs and *P* values from a multivariate analysis of the total sample.

In a series of univariate interaction analyses of each parameter × group, there was a significant interaction effect of being part-time employed (β=–5.83; SE 2.68; *P*=.03) or unemployed (β=–7.59; SE 2.52; *P*=.004), with both favoring B-CBT.

**Table 5.** Predictor analysis.

| Variables | Estimate | SE[a] | *P* value |
|---|---|---|---|
| (Intercept) | 16.85 | 4.11 | <.001 |
| Time | –0.02 | 0.00 | <.001 |
| Age | 0.00 | 0.07 | .98 |
| Female sex | –0.51 | 0.07 | .71 |
| **Marital status** | | | |
| Divorced | –0.98 | 2.07 | .64 |
| Cohabiting | 0.23 | 1.39 | .87 |
| Married | –4.08 | 2.06 | .06 |
| No answer | –1.41 | 3.98 | .73 |
| **Highest education** | | | |
| Further education 3-4 years | –1.76 | 1.45 | .23 |
| Higher education > 4 years | –3.08 | 1.76 | .09 |
| Fundamental school 9-10 years | –2.85 | 3.14 | .37 |
| High school (3 years) | –3.41 | 1.76 | .06 |
| Skilled worker | –0.80 | 2.55 | .76 |
| **Employment status** | | | |
| Part-time employed | 1.09 | 1.70 | .53 |
| Sick leave | 3.96 | 1.54 | .02[b] |
| Leave of absence | 1.48 | 3.12 | .64 |
| Retired | –0.53 | 3.48 | .88 |
| Unemployed | 0.26 | 1.63 | .87 |
| **Preference and expectancy** | | | |
| Blended care | –3.25 | 1.53 | .04[b] |
| Face-to-face | –1.18 | 1.28 | .36 |
| Credibility | 0.24 | 0.31 | .45 |
| Expectancy | –0.37 | 0.28 | .20 |
| **Usability** | | | |
| System usability | –0.05 | 0.11 | .73 |

[a] SE: standard error.

[b] *P*<.05.

### *Completion*

In total, 53 (70%) completed the treatment; 27 (71%) from the FtF-CBT group and 26 (68%) from the B-CBT group. Completers as well as noncompleters showed a significant effect of time (completers: β=–.03, *P*<.001 and noncompleters: β=–.03, *P*<.001). In an analysis of the total sample including a binary interaction term for completion, no significant interaction was seen (β=.00, *P*=.43), which indicated that there was no difference in effect between completers and noncompleters.

In the FtF-CBT group, a mean of 9.8 sessions was completed. In the B-CBT group, a mean of 9.2 sessions was completed. The mean difference was not significant ($t_{74}$=–0.70, *P*=.49). Table 6 presents the reasons for noncompletion.

We did not find any variables that significantly predicted noncompletion in multivariate analyses of the total sample nor did we find any interaction effect between any of the baseline characteristics and groups in a series of univariate analyses, indicating no difference in risk of noncompletion on any baseline characteristic between groups.

**Table 6.** Reasons for noncompletion (n=15).

| Reasons for noncompletion | Value, *n* |
|---|---|
| Inactive | 2 |
| No reason given | 7 |
| Felt it was too strenuous | 1 |
| Referred to other treatment | 2 |
| Wished to end the treatment | 2 |
| Felt unable to profit from the treatment | 1 |

## Discussion

### Principal Findings

The main aim of this study was to compare the clinical effectiveness of B-CBT with traditional FtF-CBT, because the blended format may hold the promise to combine advantages of the traditional and the new format of delivery. In this study, we found very similar trajectories of improvement in both groups as well as on measures of other parameters, such as working alliance and retention. However, it was possible to detect a significant difference between groups in slight favor of FtF-CBT.

The sample corresponded well with what is seen among patients with depression in the primary health care sector in Denmark regarding gender and age distribution [3,62]. Further, the distribution of the highest education level resembles that of the general Danish population [63].

The mean symptomatic change observed in the B-CBT condition closely approximated that of the FtF-CBT group. This is in line with meta-analyses of guided iCBT for depression [16,64] and exceeded what has been observed at the clinic "Internetpsykiatrien," from which this study recruited [17]. It also aligned well with the large effect ($d$=1.08) seen in the study of B-CBT by Lungu et al [33]. Nonetheless, an interaction effect could be detected between the 2 groups favoring FtF-CBT, although the effect was very small. This effect seems to be driven by the FtF group experiencing a faster symptom reduction during treatment, an effect that disappeared at 6 months' follow-up. However, at 12 months' follow-up, the FtF group showed a larger reduction compared with B-CBT. It is important to note, though, that none of these differences were statistically significant, thus a difference between groups cannot be inferred. Unfortunately, with the variance observed, it is not possible to infer noninferiority either (a difference no bigger than $d$=.02), due to the CIs stretching beyond the noninferiority margin. Consequently, although promising, the study is inconclusive regarding noninferiority but may support no superiority of either treatment.

The working alliance between the patient and clinician has often been argued to be one of the most important nonspecific factors of psychotherapy [65]. It is, therefore, very interesting that although half of the sessions in B-CBT were computerized, the therapeutic alliance was rated equally well in both groups. There was a tendency among the clinicians to rate it higher in the FtF group, but the difference was not significant. Similar findings are also emerging in other studies of blended care [66]. Furthermore, these studies are starting to point to details differing between B-CBT and FtF-CBT, for example, therapist ratings sometimes correlating more with treatment outcome than patient ratings conflicting with research on the working alliance in FtF-CBT [65]. Further study detailing the dynamics of the therapeutic alliance in B-CBT and the difference between that and FtF-CBT is needed.

Acceptability of the blended format seemed to be high as judged by levels of client satisfaction and working alliance, where no significant differences were observed. Furthermore, high retention rates among participants and high treatment fidelity rates for the clinicians indicated satisfaction and acceptability with the treatment.

As is commonly found [67-69], a negative correlation between intercept and slope was seen, indicating that a higher baseline severity of depression was associated with larger symptom reduction.

Interestingly, 1 variable was able to distinguish between the 2 groups in predicting outcome differently. An interaction effect was observed between employment status and group. Being part-time employed or unemployed both favored the blended care group. We speculate that this may be due to the B-CBT treatment always being available, possibly increasing the chance of treatment engagement if the participant has more free time. Consequently, this is a potential candidate variable for stratification of treatment or a prescriptive variable. In a different approach comparing variables predicting outcomes separately for the 2 groups in a larger sample across 4 countries [70], a lower quality of life and being widowed predicted lower treatment outcome in the blended condition. However, this approach does not include the parameters in a single model comparing the conditions directly in an interaction term. Nonetheless, these findings are encouraging and prompt the need to further study potential variables for stratification of patients.

This study is well aligned with previous observations of completion rates in both guided iCBT and traditional CBT [71-74]. The reasons provided by either therapists or patients for treatment dropout varied and there were too few to differentiate between the groups.

In disagreement with what has previously been found in guided iCBT, we observed that no baseline characteristics predicted noncompletion [25,75,76]. It may be speculated to be caused by the increased therapist contact in the blended format, which may serve as a protective factor against noncompletion.

## Limitations

This study compared directly the formats of delivery with a minimum of the variance explained by differences in therapeutic methods, which is both a strength and a weakness of the design. While it lends itself well to compare the 2 treatment formats, it also somewhat limits the ecological validity, making it more difficult to make inferences about the clinical effect in routine care. Furthermore, because the study recruited from Internetpsykiatrien, which offers self-referral, even though the clinic is situated in secondary care, it can be difficult to generalize to future implementations. Additionally, only the B-CBT group received reminders about homework assignments. This might be a confounder, for example, there is a risk participants in this group grew weary of the reminders, thus affecting the perception of the treatment negatively. Finally, due to the small sample size, we had difficulty inferring noninferiority, although the many observations and advanced statistical procedures appear to have compensated for that to some degree. The large EU study E-COMPARED will be able to pool data from many studies, including this one, and may thus be able to reach more robust conclusions about noninferiority.

## Conclusions

In this study, feasibility of B-CBT was demonstrated as well as large and significant within-group effect sizes were produced. In fact, it was seen that practically without loss of treatment effect, completion rates, and therapeutic alliance, it was possible to substitute half of the FtF consultations with online modules when treating adult depression. This is remarkable and lends support to the hypothesis of noninferiority of B-CBT and should lead to the further study of this promising treatment format. However, it should also be noted that small differences were observed favoring the FtF-CBT group. Although not significant, it may be that FtF treatment works faster, and has a better long-term effect for some patients. The results, therefore, need to be replicated in larger samples or with pooled data from multicenter trials as will be done in the E-COMPARED study. Additionally, further studies should explore the applicability of B-CBT in different patient populations and clinical settings. Furthermore, participants' digital health literacy should be measured in future studies.

## Acknowledgments

Funding was granted from the Research Fund of the Mental Health Services of Southern Denmark and from the Innovation Fund Denmark as part of the project ENTER (ID: 5159-00002B). Both are public funds. None of the funds have had any role in the design of the study nor in collection of data, analysis of data, interpretation of data, or writing of the manuscript. The authors acknowledge the E-COMPARED project. We also acknowledge Odense Patient Data Exploratory Network and the Mental Health Services of Southern Denmark and Centre for Telepsychiatry.

## Authors' Contributions

KM is the main author of this work and has taken part in designing and planning the study, performing analyses and writing up the article. TEA is the cosupervisor of this study and contributed to study design and statistical analysis plan. AK is the author of the E-COMPARED generic protocol, which formed the foundation for this protocol. HR is the comain supervisor, coauthor of the E-COMPARED generic protocol, and contributed to the design of this study. KKR is the main supervisor, contributed to the study design, and assisted in the process of authoring this article in its entirety. MBL and LHE have contributed to planning the study and MBL has also participated in data collection. All contributors have, in addition to the above, proofread the article.

## Conflicts of Interest

The software NoDep was developed jointly by the Region of Southern Denmark and the private partner Context Consulting. KM participated in the development process as representative of the public partner Region of Southern Denmark. KM has no affiliations to any of the private vendors involved in the project including Context Consulting and gains no economic income from sales of licenses. The other authors declare that they have no competing interests.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 496 KB - jmir_v24i9e36577_app1.pdf ]

## References

1. Mueller T, Leon A, Keller M, Solomon D, Endicott J, Coryell W, et al. Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. Am J Psychiatry 1999 Jul;156(7):1000-1006 [FREE Full text] [doi: 10.1176/ajp.156.7.1000] [Medline: 10401442]
2. Kessing LV, Hansen M, Andersen P, Angst J. The predictive effect of episodes on the risk of recurrence in depressive and bipolar disorders - a life-long perspective. Acta Psychiatr Scand 2004 May;109(5):339-344 [FREE Full text] [doi: 10.1046/j.1600-0447.2003.00266.x] [Medline: 15049770]

3. Olsen L, Mortensen E, Bech P. Prevalence of major depression and stress indicators in the Danish general population. Acta Psychiatr Scand 2004 Feb;109(2):96-103 [FREE Full text] [doi: 10.1046/j.0001-690x.2003.00231.x] [Medline: 14725589]
4. Kessler R, Chiu W, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
5. Alonso J, Lépine JP, ESEMeD/MHEDEA 2000 Scientific Committee. Overview of key data from the European Study of the Epidemiology of Mental Disorders (ESEMeD). J Clin Psychiatry 2007;68 Suppl 2:3-9. [Medline: 17288501]
6. World Health Organization (WHO). The Global Burden of Disease: 2004 Update. Geneva, Switzerland: WHO; 2004. URL: http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html [accessed 2022-08-27]
7. Huhn M, Tardy M, Spineli LM, Kissling W, Förstl H, Pitschel-Walz G, et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. JAMA Psychiatry 2014 Jun 01;71(6):706-715. [doi: 10.1001/jamapsychiatry.2014.112] [Medline: 24789675]
8. Alonso J, Codony M, Kovess V, Angermeyer M, Katz S, Haro J, et al. Population level of unmet need for mental healthcare in Europe. Br J Psychiatry 2007 Apr;190:299-306 [FREE Full text] [doi: 10.1192/bjp.bp.106.022004] [Medline: 17401035]
9. Barak A, Hen L, Boniel-Nissim M, Shapira N. A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. Journal of Technology in Human Services 2008 Jul 03;26(2-4):109-160 [FREE Full text] [doi: 10.1080/15228830802094429]
10. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
11. 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 [FREE Full text] [doi: 10.1586/erp.12.67] [Medline: 23252357]
12. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
13. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct 01;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
14. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014 May 20;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
15. Cuijpers P, Riper H. Intervenciones basadas en internet para los trastornos depresivos: Una visión general. RPPC 2015 Jan 15;19(3):209 [FREE Full text] [doi: 10.5944/rppc.vol.19.num.3.2014.13902]
16. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
17. Mathiasen K, Riper H, Andersen T, Roessler K. Guided Internet-Based Cognitive Behavioral Therapy for Adult Depression and Anxiety in Routine Secondary Care: Observational Study. J Med Internet Res 2018 Nov 28;20(11):e10927 [FREE Full text] [doi: 10.2196/10927] [Medline: 30487118]
18. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. JAMA Psychiatry 2017 Apr 01;74(4):351-359 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
19. GELLATLY J, BOWER P, HENNESSY S, RICHARDS D, GILBODY S, LOVELL K. What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. Psychol. Med 2007 Feb 19;37(9):1217-1228. [doi: 10.1017/s0033291707000062]
20. Andersson G, Carlbring P, Berger T, Almlöv J, Cuijpers P. What Makes Internet Therapy Work? Cognitive Behaviour Therapy 2009 Jan;38(sup1):55-60. [doi: 10.1080/16506070902916400] [Medline: 19675956]
21. Knowles S, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323 [FREE Full text] [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
22. Rozental A, Boettcher J, Andersson G, Schmidt B, Carlbring P. Negative effects of internet interventions: a qualitative content analysis of patients' experiences with treatments delivered online. Cogn Behav Ther 2015 Feb 23;44(3):223-236. [doi: 10.1080/16506073.2015.1008033] [Medline: 25705924]
23. Mathiasen K, Riper H, Ehlers LH, Valentin JB, Rosenberg NK. Internet-based CBT for social phobia and panic disorder in a specialised anxiety clinic in routine care: Results of a pilot randomised controlled trial. Internet Interv 2016 May;4:92-98 [FREE Full text] [doi: 10.1016/j.invent.2016.03.001] [Medline: 30135794]

24. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]
25. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol. Med 2008 Sep 24;39(5):705-712 [FREE Full text] [doi: 10.1017/s0033291708004224]
26. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED consortium. Attitudes towards digital treatment for depression: A European stakeholder survey. Internet Interv 2017 Jun;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
27. Rogers EM. Diffusion of Innovations (5th edition). New York, NY: Free Press; 2003.
28. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered internet psychotherapy for depression in primary care: a randomised controlled trial. The Lancet 2009 Aug;374(9690):628-634. [doi: 10.1016/s0140-6736(09)61257-5]
29. 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 05;15(8):e153 [FREE Full text] [doi: 10.2196/jmir.2714] [Medline: 23916965]
30. Volker D, Zijlstra-Vlasveld MC, Anema JR, Beekman AT, Brouwers EP, Emons WH, et al. Effectiveness of a blended web-based intervention on return to work for sick-listed employees with common mental disorders: results of a cluster randomized controlled trial. J Med Internet Res 2015 May 13;17(5):e116 [FREE Full text] [doi: 10.2196/jmir.4097] [Medline: 25972279]
31. Kooistra L, Ruwaard J, Wiersma J, van Oppen P, van der Vaart R, van Gemert-Pijnen JEWC, 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]
32. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013 Nov 07;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]
33. 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]
34. 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]
35. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194 [FREE Full text] [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
36. Mathiasen K, Andersen TE, Riper H, Kleiboer AAM, Roessler KK. Blended CBT versus face-to-face CBT: a randomised non-inferiority trial. BMC Psychiatry 2016 Dec 05;16(1):432 [FREE Full text] [doi: 10.1186/s12888-016-1140-y] [Medline: 27919234]
37. American Psychiatric Association. Diagnostic and Statistical Manual of Metal Disorders (4th ed. TR). Washington, D.C: Amercian Psychiatric Association; 2000.
38. Sheehan DV, Lecrubier Y, Sheehan K, Amorim P, Janavs J, Weiller E. The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33. [doi: 10.1037/t18597-000]
39. Kroenke K, Spitzer R, Williams J. 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]
40. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359 [FREE Full text] [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
41. Folker A, Mathiasen K, Lauridsen S, Stenderup E, Dozeman E, Folker M. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: A comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018 Mar;11:60-70 [FREE Full text] [doi: 10.1016/j.invent.2018.02.001] [Medline: 30135761]
42. Folker M, Mathiasen K, Lauridsen SM, Stenderup E, Dozeman E, Folker MP. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: A comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018 Mar;11:60-70 [FREE Full text] [doi: 10.1016/j.invent.2018.02.001] [Medline: 30135761]
43. Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. Journal of Consulting and Clinical Psychology 1993;61(4):620-630. [doi: 10.1037/0022-006x.61.4.620]

44. Luborsky L, Woody GE, McLellan AT, O'Brien CP, Rosenzweig J. Can independent judges recognize different psychotherapies? An experience with manual-guided therapies. Journal of Consulting and Clinical Psychology 1982 Feb;50(1):49-62. [doi: 10.1037/0022-006x.50.1.49]
45. Odense Patient data Explorative Network (OPEN). Velkommen til OPEN - Odense Patient data Explorative Network. Odense Patient data Explorative Network (OPEN). URL: https://open.rsyd.dk/ [accessed 2018-09-24]
46. Wittkampf K, Naeije L, Schene A, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29(5):388-395 [FREE Full text] [doi: 10.1016/j.genhosppsych.2007.06.004] [Medline: 17888804]
47. Rush AJ, Trivedi M, Ibrahim H, Carmody T, Arnow B, Klein D, et al. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biological Psychiatry 2003 Sep;54(5):573-583 [FREE Full text] [doi: 10.1016/s0006-3223(02)01866-8]
48. Rush A, Gullion C, Basco M, Jarrett R, Trivedi M. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996 May;26(3):477-486 [FREE Full text] [doi: 10.1017/s0033291700035558] [Medline: 8733206]
49. Nguyen TD, Attkisson C, Stegner BL. Assessment of patient satisfaction: Development and refinement of a Service Evaluation Questionnaire. Evaluation and Program Planning 1983 Jan;6(3-4):299-313. [doi: 10.1016/0149-7189(83)90010-1]
50. Attkisson C, Greenfield T. The Client Satisfaction Questionnaire (CSQ) Scales. In: Sederer LL, Dickey B, editors. Outcome Assessment in Clinical Practice. Philadelphia, PA: Williams & Wilkins; 1996:120-127.
51. Devilly G, Borkovec T. Psychometric properties of the credibility/expectancy questionnaire. Journal of Behavior Therapy and Experimental Psychiatry 2000 Jun;31(2):73-86 [FREE Full text] [doi: 10.1016/s0005-7916(00)00012-4]
52. Horvath A, Greenberg L. Development and validation of the Working Alliance Inventory. Journal of Counseling Psychology 1989 Apr;36(2):223-233 [FREE Full text] [doi: 10.1037/0022-0167.36.2.223]
53. Hatcher R, Gillaspy J. Development and validation of a revised short version of the working alliance inventory. Psychotherapy Research 2006 Jan;16(1):12-25 [FREE Full text] [doi: 10.1080/10503300500352500]
54. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
55. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J Mem Lang 2013 Apr;68(3):255-278 [FREE Full text] [doi: 10.1016/j.jml.2012.11.001] [Medline: 24403724]
56. R Core Team. R Foundation for Statistical Computing. R: A Language and Environment for Statistical Computing. R Foundation/R Project. 2018. URL: https://www.r-project.org/ [accessed 2022-08-27]
57. Kuznetsova A, Brockhoff P, Christensen R. Package: Tests in Linear Mixed Effects Models. J. Stat. Soft 2017;82(13):e1 [FREE Full text] [doi: 10.18637/jss.v082.i13]
58. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using. J. Stat. Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
59. Sawilowsky S. New Effect Size Rules of Thumb. J. Mod. App. Stat. Meth 2009 Nov 01;8(2):597-599 [FREE Full text] [doi: 10.22237/jmasm/1257035100]
60. Cohen J. A power primer. Psychological Bulletin 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
61. Dunlap W, Cortina J, Vaslow J, Burke M. Meta-analysis of experiments with matched groups or repeated measures designs. Psychological Methods 1996 Jun;1(2):170-177 [FREE Full text] [doi: 10.1037/1082-989x.1.2.170]
62. Christensen Sven Bundgaard Per Bech O. Prevalence of clinical (major) depression in general practice using the DSM-IV version of PRIME-MD. Int J Psychiatry Clin Pract 2001 Jul 12;5(1):49-54. [doi: 10.1080/136515001300225196] [Medline: 24936998]
63. Danmarks Statistik. Befolkningens højst fuldførte uddannelse. Danmarks Statistik. URL: https://dst.dk/da/Statistik/emner/uddannelse-og-viden/befolkningens-uddannelsesstatus/befolkningens-hoejst-fuldfoerte-uddannelse [accessed 2018-10-30]
64. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18 [FREE Full text] [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
65. Horvath A, Symonds B. Relation between working alliance and outcome in psychotherapy: A meta-analysis. Journal of Counseling Psychology 1991 Apr;38(2):139-149 [FREE Full text] [doi: 10.1037/0022-0167.38.2.139]
66. Preschl B, Maercker A, Wagner B. The working alliance in a randomized controlled trial comparing online with face-to-face cognitive-behavioral therapy for depression. BMC Psychiatry 2011 Dec 06;11(1):189. [doi: 10.1186/1471-244x-11-189]
67. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013 Feb 26;346(feb26 2):f540 [FREE Full text] [doi: 10.1136/bmj.f540] [Medline: 23444423]
68. El Alaoui S, Ljótsson B, Hedman E, Kaldo V, Andersson E, Rück C, et al. Predictors of Symptomatic Change and Adherence in Internet-Based Cognitive Behaviour Therapy for Social Anxiety Disorder in Routine Psychiatric Care. PLoS One 2015;10(4):e0124258 [FREE Full text] [doi: 10.1371/journal.pone.0124258] [Medline: 25893687]

69. Warmerdam L, Van Straten A, Twisk J, Cuijpers P. Predicting outcome of Internet-based treatment for depressive symptoms. Psychother Res 2013 Sep;23(5):559-567. [doi: 10.1080/10503307.2013.807377] [Medline: 23848944]
70. Friedl N, Krieger T, Chevreul K, Hazo JB, Holtzmann J, Hoogendoorn M, et al. Using the Personalized Advantage Index for Individual Treatment Allocation to Blended Treatment or Treatment as Usual for Depression in Secondary Care. J Clin Med 2020 Feb 11;9(2):490 [FREE Full text] [doi: 10.3390/jcm9020490] [Medline: 32054084]
71. El Alaoui S, Ljótsson B, Hedman E, Svanborg C, Kaldo V, Lindefors N. Predicting Outcome in Internet-Based Cognitive Behaviour Therapy for Major Depression: A Large Cohort Study of Adult Patients in Routine Psychiatric Care. PLoS One 2016 Sep 12;11(9):e0161191 [FREE Full text] [doi: 10.1371/journal.pone.0161191] [Medline: 27618548]
72. Hadjistavropoulos HD, Pugh NE, Hesser H, Andersson G. Predicting Response to Therapist-Assisted Internet-Delivered Cognitive Behavior Therapy for Depression or Anxiety Within an Open Dissemination Trial. Behav Ther 2016 Mar;47(2):155-165. [doi: 10.1016/j.beth.2015.10.006] [Medline: 26956649]
73. Nordgreen T, Havik OE, Ost LG, Furmark T, Carlbring P, Andersson G. Outcome predictors in guided and unguided self-help for social anxiety disorder. Behav Res Ther 2012 Jan;50(1):13-21. [doi: 10.1016/j.brat.2011.10.009] [Medline: 22134140]
74. Wierzbicki M, Pekarik G. A meta-analysis of psychotherapy dropout. Professional Psychology: Research and Practice 1993 May;24(2):190-195 [FREE Full text] [doi: 10.1037/0735-7028.24.2.190]
75. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an ‘individual patient data’ meta-analysis. Psychol. Med 2015 Apr 17;45(13):2717-2726. [doi: 10.1017/s0033291715000665]
76. Christensen H, Griffiths K, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]

---

## Abbreviations

**B-CBT:** blended cognitive behavioral therapy
**BDI-II:** Beck Depression Inventory II
**CBT:** cognitive behavioral therapy
**CEQ:** Credibility and Expectancy Questionnaire
**CSQ-8:** 8-item Client Satisfaction Questionnaire
**DSM-IV-TR:** Diagnostic and Statistical Manual of Mental disorders 4th edition text revision
**ENTER:** e-Mental Health Research
**EU:** European Union
**FtF-CBT:** face-to-face cognitive behavioral therapy
**M.I.N.I.:** Mini-International Neuropsychiatric Interview
**OPEN:** Odense Patient data Exploratory Network
**PHQ-9:** 9-item Patient Health Questionnaire
**QIDS-16-SR:** 16-item Quick Inventory of Depressive Symptomatology Self-Report
**SE:** standard error
**TAI:** Technical Alliance Inventory
**TAU:** treatment as usual
**WAI:** Working Alliance Inventory-Short Revised
**WHO:** World Health Organization

---

*Edited by R Kukafka; submitted 18.01.22; peer-reviewed by C Tsai, M Wu; comments to author 10.02.22; revised version received 02.03.22; accepted 15.06.22; published 07.09.22.*

*Please cite as:*
*Mathiasen K, Andersen TE, Lichtenstein MB, Ehlers LH, Riper H, Kleiboer A, Roessler KK*
*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;24(9):e36577*
*URL: https://www.jmir.org/2022/9/e36577*
*doi:10.2196/36577*
*PMID:36069798*

Original Paper

# The Value of Technology to Support Dyadic Caregiving for Individuals Living With Heart Failure: Qualitative Descriptive Study

Noor El-Dassouki[1], BSc, MSc; Kaylen Pfisterer[1], BSc, MSc, PhD; Camila Benmessaoud[1], BSc, MHI; Karen Young[1,2], BCom; Kelly Ge[1,2], BHS; Raima Lohani[1], BScN, MHI; Ashish Saragadam[1,3]; Quynh Pham[1,2,4], PhD

[1]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Health Sciences, Faculty of Health, University of Waterloo, Waterloo, ON, Canada

[4]Telfer School of Management, University of Ottawa, Ottawa, ON, Canada

**Corresponding Author:**
Quynh Pham, PhD
Centre for Global eHealth Innovation
Techna Institute
University Health Network
Toronto General Hospital R. Fraser Elliott Building, 4th Floor
190 Elizabeth Street
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 4800 ext 4765
Email: q.pham@uhn.ca

## Abstract

**Background:** The demand for health services to meet the chronic health needs of the aging population is significant and remains unmet because of the limited supply of clinical resources. Specifically, in managing heart failure (HF), digital health sought to address this gap during the COVID-19 pandemic but highlighted an access issue for those who could not use technology-mediated health care services without the support of their informal caregivers (ICs). The complexity of managing HF symptoms and recurrent exacerbations requires many patients to comanage their illness with their ICs in a care dyad, working together to optimize patient outcomes and health-related quality of life. However, most HF programs have missed the opportunity to consider the dyadic perspective despite interdependencies on HF outcomes.

**Objective:** This study aims to characterize the value of technology in supporting caregiving for individuals living with HF.

**Methods:** Motivated by an observed unique pattern of engagement in patients enrolled in our Medly HF management program at the Peter Munk Cardiac Centre in Toronto, Canada, we conducted 20 semistructured interviews with a convenience sample of ICs. All interviews were analyzed using the iterative refinement of a codeveloped codebook. The team maintained reflexivity journals to reflect the impact of their positionality on their coding. Themes were first derived deductively using HF typologies (patient-oriented dyads, caregiver-oriented dyads, and collaboratively oriented dyads) and then inductively refined and recategorized based on concepts from the van Houtven et al framework.

**Results:** We believe that there is a need to formally and intentionally expand HF technologies to include dyadic needs and goals. We suggest defining 3 opportunities in which value can be added to technological design. First, identify how technology may be leveraged to increase psychological bandwidth by reducing uncertainty and providing peace of mind. We found that actionable feedback was highly desired by both partners. Second, develop technology that can serve as a member of the dyad's support system. In our experience, automated prompts for patients to take measurements can mimic the support typically provided by ICs and ease their workload. Third, consider how technology can mitigate the dyad's clinical knowledge requirements and learning curve. Our approach includes real-time actionable feedback paired with a human-in-the-loop, nurse-led model of care.

**Conclusions:** Our findings identified a need to focus on improving the dyadic experience as a whole by building IC functionality into digital health self-management interventions. Through a shared model of care that supports the role of the patient in their own HF management, includes ICs to expand and enhance the patient's capacity to care, and acknowledges the need of ICs to care for themselves, we anticipate improved outcomes for both partners.

***(J Med Internet Res 2022;24(9):e40108)*** doi:10.2196/40108

**KEYWORDS**

heart failure; digital therapeutics; remote patient management; caregiving; dyadic management

# Introduction

## Background

The demand for health services to meet the chronic health needs of the aging population is significant but remains unmet because of the limited supply of clinical resources. Although the pandemic-induced scale and spread of digital health sought to address this gap and widen access to care [1], it also highlighted a portion of the population who could not use or benefit from technology-mediated health care services without the support of their informal caregivers (ICs). Defined as individuals who provide unpaid care and assistance to friends or family members because of a health condition, 2.7 million Canadians aged ≥45 years identify as ICs and provide approximately 75% of home care services to support aging in place [1,2]. In this context, we use the general term ICs to refer to either primary or secondary ICs to account for the intricacies of family-oriented care and the ecosystem that the broader circle of care provides within the context of a patient's care community.

Within digital health and the broader health care system, having support from an IC is particularly important for individuals living with heart failure (HF). As the most complex and costly chronic condition to manage in health care, HF is the single most common reason for hospital admissions and readmissions in Canada [3]. As a result, patients require constant care, largely because of the complexity of HF symptom management and recurring exacerbations [4,5]. This leads many individuals living with HF to comanage their illness with their ICs in a dyadic arrangement, with dyads working together to optimize patient outcomes and health-related quality of life.

Although substantial work has been conducted on patient self-care interventions, most HF programs have missed the opportunity to consider the broader perspective of the dyad [6,7]. As each member of the dyad influences the behavior and well-being of the other, integrating dyadic considerations into the design of HF interventions has been shown to enhance patient health outcomes [6-11]. Most existing HF interventions geared toward dyads have emphasized peer support and mentoring programs, patient education programs, face-to-face education and visits, brochures, and voice calls [7,12]. Although largely unexplored, we anticipate technology to be well positioned to help support dyads to (1) perform several complex tasks associated with the daily demands of HF, such as planning appointments and managing medications; (2) receive real-time feedback as a supplementary support system; and (3) access tailored educational resources to address the high degree of caregiving uncertainty experienced because of insufficient knowledge and training.

We were motivated to pursue this research as our Medly clinical and implementation team observed a unique pattern of engagement in certain patients enrolled in our Medly HF management program at the Peter Munk Cardiac Centre in Toronto, Canada [13]. Patients within the program use the Medly platform as the standard of care to capture daily health measures, including weight, blood pressure, heart rate, and HF-related symptoms. The embedded Medly algorithm immediately analyzes the entered readings against set personalized thresholds and provides the patient with instant feedback and instructions. The algorithm also alerts the patient's care team on the Medly dashboard for further assessment and triage [14]. Owing to the unacceptably high risk to patients living with HF coming into the hospital at the height of the pandemic, cardiologists in the program began enrolling patients into Medly who would typically not be eligible for the program. The team observed that this new cohort of patients had low digital health literacy and English language proficiency but could be enrolled as they were onboarded with an IC who had agreed to use Medly on their behalf. This off-label dyadic use of an HF digital therapeutic (DTx) allowed us to further explore how we could provide better support for dyads to engage in effective shared care through technology.

## Objective

This study sought to characterize how technology, specifically DTx, might support ICs engaged in the dyadic management of HF to optimize their caregiving practices and improve shared outcomes. We define DTx as the "evidence-based, clinically evaluated software [used] to treat, manage, and prevent a broad spectrum of diseases and disorders" [15]. Specifically, we sought to answer two research questions: (1) what are the challenges of caregiving for an individual living with HF, and (2) how can technology support the dyadic management of HF to enable improved outcomes for both partners?

# Methods

## Conceptual Frameworks

To structure our research strategy and approach, we began by conducting an environmental scan of the existing literature on the dyadic management of HF. We also sought a framework that could help examine the relationship between dyad type and the value of technology in HF management. Buck et al [16,17] conceptualized an HF care dyadic typology by integrating and applying 2 theoretical models, the actor-partner interdependence model and interdependence theory, to the dyadic management of HF. They described 4 categories of dyadic relationships, which we collapsed into 3 to align with our approach: patient-oriented dyads, caregiver-oriented dyads, and collaboratively oriented dyads. In patient-oriented and caregiver-oriented dyads, the person living with HF or the IC takes on most of the self-care responsibilities of the patient living with HF. Conversely, in collaboratively oriented dyads, both the patient and IC have joint involvement in HF self-care [16]. This framework allowed us to strengthen the theoretical basis of our work, add to the evidence base in the validation of this model, and advance our understanding of the role that dyad type plays in the use of technology in HF management. We continually returned to this framework during the course of data collection and analysis, reflecting on the ways in which our data were bounded by the typology and the use of the dyadic relationship as the main frame of analysis.

Upon completing the initial analysis of our data using the typologies outlined by Buck et al [16,17], we sought an organizing framework to further consider and optimize future evaluation efforts. The van Houtven et al [18] framework considers how the baseline characteristics of the patient-caregiver dyad influence the way an intervention is designed and how the caregiver engages with it. The efficacy of an intervention is then evaluated based on the changes it contributes to (1) caregiver activities and (2) caregiver and patient outcomes, which include psychological and physical health, health care use, and financial status. Although this study aimed to create rather than evaluate an intervention, we felt that the work by van Houtven et al [18] provided a strong, standardized framework for determining which factors need to be considered when designing an effective intervention that optimizes patient and caregiver outcomes.

## Study Design

### *Overview*

This qualitative descriptive study was conducted at the University Health Network (UHN) between April 2021 and March 2022 during the COVID-19 pandemic. All recruitment and data collection activities were conducted virtually. Data collection was conducted by authors NE and CB, and the data analysis team comprised research staff (NE and CB) and research students (KY and KG).

### *Positionality of the Data Analysis Team*

The term positionality refers to the ways in which individuals identify and relate to various social dimensions (ie, gender, race, and ethnicity) [19]. A key element of high-quality qualitative research is sincerity, achieved through self-reflexivity and transparency on the part of the researcher [20]. Being able to openly recognize the researcher's positionality, experiences, and assumptions and center the active role that the researcher plays in the interpretation of data is a central tenet of reflexive thematic analysis [21]. In alignment with the related concept of researcher subjectivity as a valuable resource, the authors maintained a reflexivity journal and described their social positionalities to understand how their backgrounds and experiences intersected with data analysis [22].

NE's experiences as a caregiver, woman of color, and health equity researcher contributed to how she interacted with and analyzed the interview data. Having conducted interviews and built a relationship with interviewees, she felt a personal connection to the data and worked to ensure that the invaluable extent of caregiver support in the context of HF was thoroughly reflected in the analysis. Her identities also drove her to achieve representation of diverse racial and ethnic groups within the study sample and balance the representation of all participants within the analysis while also centering the stories, experiences, and suggestions of those who are often underrepresented in health and caregiving research.

CB's social positionality as a young adult, first-generation Canadian, and member of an ethnocultural minority group largely influenced her ability to empathize with, and understand, the caregiver experience. Although her deeply rooted and culturally embedded moral values surrounding shared familial care allowed her to relate to the motivations behind providing care to a loved one, CB's lack of firsthand experience in the caregiving space posed a unique challenge to understanding the true extent of daily caregiving and its unique physical, mental, and emotional demands. As an interviewer, CB found that having vulnerable conversations with the participants allowed her to subconsciously build a personal connection and empathize more effectively with the participants and their lived experiences. The personal bond, combined with the emotional recollection experienced by CB, led to unique opinions and perspectives while conducting data analysis.

KY is a second-generation Canadian settler and a woman of color. She has acted as a caregiver, translator, and health care intermediary for multiple family members with chronic illnesses. During the analysis, these identities and labels led her to be conscious and emotionally responsive to how her experiences intersected with each participant's caregiving story, especially regarding the cultural context of caregiving. She did not complete any interviews during the data collection phase and was only familiar with the participants to the extent to which they provided personal information in each interview transcript.

KG's positionality as a second-generation Canadian and a woman of color informed her analysis. Although she has not acted directly as a caregiver, her experiences with and observations of the challenges of navigating health care and digital health tools supported her understanding and connections to the participants. Her experiences with equity-focused digital health research and university coursework in the social determinants of health also led to heightened attention to the limitations of technology in the context of complex health systems. She did not conduct interviews during data collection but contributed to the interpretive analysis of transcripts.

## Sample

We obtained a convenience sample of adult ICs (aged ≥18 years) of patients enrolled in the Medly HF management program at UHN. First, the research coordinators reviewed the Medly patient list and reached out to the patient care team for additional screening to identify those who had or may have had ICs engaged in their care. An email was then sent to the Medly patients by the research coordinators, introducing the study and asking for a referral to connect with the IC. A total of 255 patients were approached. All ICs referred by their patient partners expressing interest in being interviewed by the research team were eligible to participate in this research.

## Data Collection

### *Overview*

We conducted 20 semistructured individual interviews with a group of ICs across various age groups, ethnic backgrounds, relationship types, and additional social dimensions, as demonstrated in the *Results* section. Our sampling strategy was guided by the exploratory nature of our research aims, pragmatic judgment, and the concept of *information power*, which suggests that the sufficiency of the sample size can be determined based on the pertinence and salience of the information shared by participants [23,24]. In the approach described in the following sections, we aimed to identify common patterns of beliefs and behaviors from participant interviews that were relevant to our research questions. These patterns painted a rich and complex picture of their experiences of caregiving for patients living with HF and how these experiences intersected and interacted with their use of Medly. Demographic information was collected from all participants on age, gender, language, ethnicity, education, living arrangement, area of residence, marital status, employment status, and annual household income. Questionnaires were completed electronically using the REDCap (Research Electronic Data Capture; Vanderbilt University) tool hosted at UHN [25,26]. Demographics are described thoroughly in the *Results* section.

### Interviews

Using a semistructured interview guide (Multimedia Appendix 1), interviewers NE and CB guided caregivers through questions that explored three main themes: (1) caregivers' relationship with their care recipient and their experience with caregiving, (2) the caregiver's role in and views on the Medly program, and (3) opportunities to improve the Medly experience to further support the dyad. Caregivers were asked to reflect on the rewards and challenges of the caregiving experience and provide insights into how they would define a better future. Interviewers guided caregivers through discussions about their personal goals and the barriers they faced in achieving them.

Owing to the COVID-19 pandemic restrictions, all interviews were conducted virtually, either using Microsoft Teams or over the phone. The interviews were audio recorded, and a professional service transcribed the recordings. As expected, the virtual setting of the research posed unique challenges, including establishing rapport with participants and addressing technical issues. Participants chose to conduct their interviews over a video call or over the phone (audio only) based on personal preference. Although phone calls were more accessible to the participants, it was more challenging to build rapport using this modality. Rapport was more effectively built during video call interviews because of the ability to use body language to communicate and create a more human-centered and less formal research environment. To address the nuances of virtual meetings and their associated challenges, our team encouraged participants to complete their sessions through video calls and started all interviews with rapport-building conversations to foster a safe space.

In both phone and video modalities, participants often encountered technical frustrations with electronic consent methods and technical issues associated with navigating Microsoft Teams. To mitigate these frustrations, our team developed a step-by-step guide on how to navigate the teleconferencing tool, which was distributed to the participants via email before their session. For those who continued to experience technical issues, a study team member was able to successfully troubleshot and provide technical support via phone.

### Codebook Development

The preliminary codebook was developed by 3 team members (NE, CB, and KG) using data from the first 6 interviews based on IC activities outlined by Buck et al [16] and the initial themes identified through familiarization with the transcripts. Specifically, the first 3 interviews were used to conduct a preliminary group listening activity with authors NE, CB, RL, and QP to discuss key ideas, thoughts, and potential feature suggestions for the family-centered model of dyadic digital health. NE, CB, and KG used the remaining 3 interviews to perform a preliminary analysis using the initial codes, with each transcript coded by 2 research team members. Iterative refinement of the codebook was continued with the aim of condensing existing codes, removing ambiguity, and establishing consensus in the codes. This process yielded the final version of the codebook for use in formal data analysis.

## Data Analysis

### Transcript Coding

We conducted a thematic content analysis of the interview data [27]. A total of 4 research team members (NE, CB, KY, and KG) used the NVivo (QSR International) software to code all the transcripts individually. First, the coders independently read each transcript to familiarize themselves with the data. The transcripts were then assigned and coded independently, while the coding team met weekly to discuss questions, resolve differences, and refine the codebook through an iterative review of the themes derived from additional interviews. The team maintained reflexivity journals to reflect on how their experiences, background, and positionality affected and intersected with their coding work. Themes were first derived deductively based on the structure of the interview guide and using HF typologies, as defined by Buck et al [16]. We then inductively refined the analysis to incorporate the identified themes, borrowing from the approach outlined by Braun and Clarke [27]. Finally, the themes identified from the transcript coding were recategorized based on concepts from the van Houtven et al [18] framework, which provides a structure for evaluating caregiver interventions.

We used a multitude of practices and methods to ensure the quality of our study, as described in the 8 *big-tent* criteria for high-quality qualitative research conceptualized by Tracy et al [20]. Specifically, we aimed to build methodological rigor through the use of relevant frameworks, rich descriptions of the context and sample from which data were collected, and the use of appropriate and theoretically informed data collection and analytical methods. Sincerity, another key criterion, is spoken to in our description of our reflexivity processes and the filter of our positionality statements; that is, we seek to uncover and make ourselves and our readers aware of our motivations for pursuing this work and honestly put forward the biases we may have held in the process of data collection and analysis.

### Statistical Analyses

Descriptive statistics are reported for all demographic data collected from the study sample. The number and percentage of participants who responded to each question, as well as the mean age of participants (reported with the SD), were calculated and are presented in the *Results* section. Missing responses were removed from the total count when calculating the percentage and mean age and were noted where applicable.

### Ethics Approval

All recruitment and data collection activities were approved by the UHN Research Ethics Board (REB 20-5238).

## Results

### Overview

A total of 20 ICs were interviewed in this study. The average age of the participants was 63 (SD 7.5) years, and most identified as women (17/20, 85%) and White (14/20, 70%). However, our sample also included 25% (5/20) of participants who spoke English as a second language and 25% (5/20) who identified with diverse racial and ethnic groups. We acknowledge that this sample experienced favorable socioeconomic conditions, with 65% (13/20) of participants educated above the high school level and 40% (8/20) earning an annual household income of >CAD $100,000 (US $77,170). Most caregivers in our sample were assessed to belong to a collaborative dyad typology (10/20, 50%; 5/20, 25% patient-oriented; 5/20, 25% caregiver-oriented). Approximately 55% (11/20) noted that they were wives caring for their husbands who lived with HF. For a detailed breakdown of the demographic information, refer to Table 1.

Interviewers and participants were able to connect through shared experiences and the ease of conversation about the interview topics. Our qualitative analysis yielded two salient themes pertaining to (1) IC experiences with HF caregiving activities and (2) the role of technology in facilitating HF caregiving activities within a dyad. Consistent with the study by van Houtven et al [18], these themes encompassed the amount of caregiving, psychological skills, support-seeking activities, and domains of clinical knowledge.

**Table 1.** Demographic characteristics of informal caregiver participants (N=20).

| Individual-level variable | Value |
|---|---|
| **Age[a] (years)** | |
| Values, n (%) | 19 (95) |
| Values, mean (SD) | 63 (7.5) |
| **Gender[b], n (%)** | |
| Woman | 17 (85) |
| Man | 3 (15) |
| **Patient-caregiver dyad typology, n (%)** | |
| Patient-oriented | 5 (25) |
| Caregiver-oriented | 5 (25) |
| Collaborative | 10 (50) |
| **Care relationship, n (%)** | |
| Wife caring for husband | 11 (55) |
| Husband caring for wife | 3 (15) |
| Daughter caring for parent | 4 (20) |
| Mother caring for daughter | 1 (5) |
| Friend caring for friend | 1 (5) |
| English as a second language, n (%) | 5 (25) |
| **Race or ethnicity, n (%)** | |
| South Asian | 1 (5) |
| Black: African or Caribbean | 2 (10) |
| Indigenous or Aboriginal | 1 (5) |
| Latin American | 1 (5) |
| White: European or North American | 14 (70) |
| Other(s) | 1 (5) |
| **Highest level of education, n (%)** | |
| High school diploma | 4 (20) |
| College trade or technical diploma | 5 (25) |
| University undergraduate degree (eg, BA or BEng) | 6 (30) |
| University professional designation (eg, MD or MBA) | 2 (10) |
| Other(s) | 3 (15) |
| **Employment status, n (%)** | |
| Retired | 9 (45) |
| Unemployed | 1 (5) |
| Working full-time or part-time | 8 (40) |
| Other | 2 (5) |
| **Annual household income (CAD $ [US $])[a], n (%)** | |
| <49,999 (<38,537) | 2 (10) |
| 50,000–100,000 (38,538–77,075) | 8 (40) |
| >100,000 (>77,075) | 8 (40) |

[a] Indicates missing responses.

[b] Participants were provided with the option to select “non-binary” or “prefer not to answer,” of which no selections were made.

## IC Experiences With HF Caregiving Activities

### Overview

A critical factor that inherently influences the amount of caregiving that an IC engages in is the nature of their dyadic typology. Our sample included patient-oriented (patient performs most tasks), collaborative (patient and caregiver share care tasks), and caregiver-oriented (caregiver performs most tasks) dyad typologies, each conferring distinct caregiving requirements. Specifically, a caregiver-oriented dyad often provides a greater amount of caregiving as the patient requires more support with activities of daily living along with improving and maintaining self-care. In addition, compared with patient-oriented and collaborative dyads, patients in caregiver-oriented relationships are often much more dependent on their caregivers:

> *I basically do all of it. I make sure her appointments are done; I take her to all her appointments...I do all her medications, I fill her pill boxes, I call her on the phone when it's time to take her medications...I do all her vitals in the morning, I do her blood pressure, her heart rate, her temperature twice a day. I look after anything health.* [C05, caregiver-oriented]

In contrast, patients in patient-oriented dyads were able to maintain their independence with most of their self-care tasks while receiving support from the caregiver when needed:

> *The other ways that I help him is I try to encourage him to go for walks or to, like, be more—a little bit more active...But other than that, he's pretty self-sufficient in doing his own [care]—he cooks and [eats] the right things.* [C08, patient-oriented]

The level of involvement of ICs in the care can also depend on the extent of the barriers that the care recipient experiences in interacting with the health care system (eg, visual impairments, cognitive implications, and language barriers) and the desire to provide their own care. In some instances, caregivers provide more care and are able to cope well with this arrangement, whereas others provide less care than desired because of the patient's preference to be self-sufficient and manage their own care:

> *My husband wants to be engaged. If he feels, he always feels left out. And his favourite comment to me is, "it's my body. It's my medical condition, and everyone is making decisions without my being involved." My husband is colourblind, so he can't see. He can't participate. He couldn't input the information [into Medly], because when it comes up [it's in] red...So this is why I'm managing it.* [C09, caregiver-oriented]

The amount of caregiving influences the overall IC experience as dyads continually adapt and navigate effective strategies to manage living with a diagnosis of HF within their unique circumstances. Participants from all typologies shared their caregiving experiences, the uncertainty embedded in the role, and the specific challenges they faced around (1) living with pervasive unpredictability and lack of control, (2) support-seeking behaviors to mediate the degree of IC burden, and (3) the high-stakes learning curve associated with HF diagnosis and management. Each of these is expanded in the following sections.

### Living With Pervasive Unpredictability and Lack of Control

Across all dyadic typologies, the sudden onset and unpredictable nature of an HF diagnosis often leave caregivers feeling helpless and reporting a lack of control over the patient's HF prognosis. In addition, the severity and impact of HF exacerbation events throughout the HF journey can be traumatic:

> *It was a scary moment having her collapse in my arms that day and then seeing them work on her for over half an hour trying to get her to come back to life. That was really scary. The doctor said to her that she had a 3% chance of surviving and the fact that she didn't have any brain damage or organ damage was another miracle.* [C17, collaborative]
>
> *For me as a caregiver, at times it can be frustrating—I guess because of the fear. The fear of what if this [cardiac arrest] happens again? I certainly don't want it to happen again, but living with heart failure, it's...you just don't know.* [C10, patient-oriented]

The challenge of uncertainty makes every day unique and difficult to anticipate for caregivers:

> *We don't know what to expect day-to-day. It's not like we can say "Well it's been a good week, you know, he'll be fine tomorrow," because tomorrow he may not be fine...You can't make long-range plans.* [C13, collaborative]

The IC's sense of self is often affected by sudden and unexpected changes in the care recipient's health arising from an HF diagnosis or exacerbation event. The lack of time to grieve and process the change results in caregivers feeling overwhelmed and helpless:

> *It took a couple of years to get over [my husband's cardiac arrest]. You're doing CPR on your husband. So it was just the trauma of going through that experience. And then when he did come home, seven weeks later, he's so weak. He ended up collapsing on the floor. It [wasn't until] a couple years later we had gone to the ICD Clinic and one of the doctors there, he said, "Well how did you feel?" That was the first time someone had asked me how I had felt through that experience. So I was ready to break down in tears at that time.* [C10, patient-oriented]

These rapid changes also affect the dynamic within the dyad and may confer additional stress on the IC as they work to balance caregiving for patients engaging in self-care activities. There is often guilt associated with the latter, as the caregiver struggles to take time away from the caregiving role to focus on their own physical health and well-being. These intersecting feelings of self-neglect and grief are especially amplified in caregiver-oriented and collaborative dyads, as caregivers are required to adjust their lifestyle and activities to a larger degree to accommodate the care recipient's needs:

> *Unfortunately, what happens is as one partner becomes less active, the second person also becomes less active because you feel guilty. For example, we could go out for a walk, which is recommended. But obviously, my husband doesn't feel the same way every day...So if he can't go out, I'm not going to go out and leave him.* [C09, caregiver-oriented]
>
> *I don't do what I used to do at all. And I don't want to sound selfish or anything but it seems like anything I need or I want to do is always second. Everything is always based around him, which is how it should be, but sometimes emotionally that gets to you where you just kind of, like, want to run away and have a date with yourself sort of thing.* [C18, collaborative]

The lack of control that the caregiver has over the patient's health behaviors, reactions, and perceptions may contribute to additional stress on the IC and add further strain on the dyad's relationship. The degree of congruence between the IC and care recipient regarding expectations of how to engage in care tasks and who is responsible for different tasks affects their appraisal of the caregiving experience. A dyad with similar expectations generally predicts a more sanguine approach to caregiving, which may reduce the psychological toll of the experience. However, discrepant views on dyadic expectations and the patient's care plan may result in additional stress for the IC, as the dyad shifts between being independently oriented, as in caregivers providing support, and a collaborative approach to care where care recipients are empowered and their independence is fostered:

> *We have this back and forth because he'll just take [his Medly readings] and leave [the paper that has the readings written on them] and then expects me to look for it and put it in [the Medly app]. And I say, "No. You [do it]' because I just like him to be a little bit on the up and up, like to being responsible in doing that."* [C08, patient-oriented]

The psychological toll associated with discrepant views is especially heightened for caregivers in caregiver-oriented and collaborative dyads, where the caregiver is more heavily involved in the patient's care.

### *Support-Seeking Behavior to Mediate the Degree of IC Burden*

Outlets that diffused stress or acted to alleviate caregiver burden were crucial components of caregivers' support systems and, by extension, the dyad as a whole. Outlets included family, friends, health care providers, financial stability, respite hours, and disease management tools among others.

Experiencing a high level of uncertainty in the caregiving journey without the presence of a support system decreased the ICs' ability to provide quality care and increased caregiver burden and burnout. This was most salient for caregiver-oriented dyads, whereby ICs were significantly relied on for support from their care recipients.

> *We don't have extended family here. Or friends. So it's just my family and that's it. Him and I and my son and that's it...I have been to the point where I can't take it anymore, overloaded, overwhelmed. And it affects me a lot.* [C20, caregiver-oriented]

Conversely, the dyadic relationship typically improved when caregivers had the ability to know when to seek support and had support available to diffuse the stress and responsibility they experienced. In this way, support systems aided the whole dyad; the care recipient continued to receive quality care, and the caregiver was able to receive respite from caregiving responsibilities:

> *So what I personally do is...one of my sisters...comes down and spends two nights with me every week...And it gives me a bit of relief, it gives me a bit of socializing that's just for me, where I can go outside and be in the bush for an hour and she stays in the house [with the patient]...So she's like a second me when she's here...And if it wasn't for her doing that I'm not sure what state I would be.* [C18, collaborative]

Finding appropriate support systems enables dyads to find a balance and simultaneously contributes to improved self-efficacy and resiliency among caregivers. Furthermore, the opportunity to share experiences with others who have had similar HF experiences can provide an invaluable source of support. It provides caregivers with reassurance, guidance, and a look into their future:

> *I was completely lost [caregiving for my daughter with HF]. And out of the blue, a friend of mine revealed to me that she had lived with heart failure for 20 years...She gave me so much hope, so much courage. That call came to me at a time when I didn't know what to expect. She used to call [my daughter to] ask her how she was feeling and give her own tips of how she had coped. So I've been lucky to have supportive friends as well as family. And that's what has really kept me going as a caregiver. And as a mother. It's not just a caregiver hired, but a caregiver because she's your child.* [C21, collaborative]

### *The High-Stakes Learning Curve Associated With HF Diagnosis and Management*

Caregivers described a high degree of uncertainty in their care experience, often derived from their fundamental lack of HF knowledge and professional training. This lack of knowledge inherently left caregivers feeling less confident in their ability to support their loved one with appropriate symptom management. Although knowledge can be a powerful tool to help reduce uncertainty, caregivers also shared that finding high-quality, trustworthy resources is difficult. Navigating where to turn and what to trust is complex, especially when sourcing evidence-based and up-to-date guidance on HF management. Self-doubt weighs heavily on these caregivers. The yearning for practical knowledge and greater direction on how to support their care recipient is a common sentiment felt among ICs who have more hands-on experience and those who are responsible for several care tasks. There are numerous opportunities to doubt their own subjective assessment:

> *He had an infection. But what's normal? What isn't? And we've heard every individual is different, but even to have the ability for me as a caregiver to say oh, there seems to be more discharge at the driveline site. What are you looking for? Or what would be a red flag for that?* [C14, collaborative]

> *Sometimes I'll just say, "Ok, well just lie down." But who knows if that's the right thing to do.* [C06, collaborative]

> *The only thing in the beginning was the medication. Because all the words are so foreign to me. I just take notes and it's up to him to take the medication...I was afraid that I would mess up on that...miss a time when he was supposed to have it.* [C13, collaborative]

Moreover, the frequency of HF exacerbations and the need for ongoing supportive care to prevent and manage them leaves ICs with a sharp, high-stakes learning curve as they navigate and manage the HF journey:

> *I have to understand or accept that fact that this is who he is right now, this is what he can do right now, this is what I can expect from him, no more. It's been a learning curve for me.* [C20, caregiver-oriented]

## The Role of Technology in Facilitating HF Caregiving Activities Within a Dyad

### *Overview*

Although Medly was designed to be used independently by the patient, many ICs took over the Medly user role on their behalf to enable participation in the program during the COVID-19 pandemic. Other dyads collaboratively engaged in the use of Medly to provide supportive shared care. Although caregivers tended to trust the medical guidance provided by Medly as it came from a trusted institution, this was not always the case for some care recipients. Dealing with nuanced perceptions of trust within the dyad posed unique challenges for the caregiver as they attempted to address the patient's hesitancy. As Medly is a prescription-based DTx, pre-existing relationships with clinical staff and the patient's health care team also contributed to a sense of trust within the program. Building on this foundation of trust fostered an environment in which Medly could serve as both a DTx and a component of the IC's support system. The remainder of this section outlines the dyads' experiences and the potential role of technology in supporting ICs by (1) reducing uncertainty and providing peace of mind, (2) acting as a *member* of the dyad's support system, and (3) mitigating the dyad's clinical knowledge requirements.

### *Technology May Reduce Uncertainty and Provide Peace of Mind*

Earlier, we noted the magnitude, impact, and weight that uncertainty imparted to ICs. Medly's regular check-ins and accessible connection to clinical staff provide a release valve to diffuse stress associated with the care recipient's unknown disease progression and acute status. In this way, technology can reduce uncertainty by providing dyads with personalized directions on what they need to do to best manage the patient's condition on a day-to-day basis:

> *When I put in the data if there's anything a little bit off with it, it'll tell me, this is too high, this is too low, this is a concern. I like that. Because then I know that it'll tell me if I should make him redo the measurements in the afternoon.* [C08, patient-oriented]

> *The quick response to the information that's inputted is fabulous. It tells me whether we're on track or we're not on track. And then I could modify on a day to day basis...it also provides me [with] a trend line...if I want to look at the last week and see what the trend is, I can see what we're doing.* [C09, caregiver-oriented]

The ability to connect with experts through Medly also provides invaluable peace of mind to the dyad:

> *Sometimes they'll just notice that [her weight is] starting to creep up and they'll [advise], "OK you should take that drug and go get the blood tests done." And so it's nice knowing that somebody is keeping an eye out.* [C17, collaborative]

Through Medly, the care recipient's health is monitored on an ongoing basis where they would otherwise be left alone to manage the condition without expert oversight outside of synchronous touchpoints with the clinical team. This peace of mind has a profound impact on relieving a significant proportion of caregiver burden:

> *I mean she came home from the hospital...her procedure was a bit critical in the hospital...Just being able to come home and hook [Medly] up, it just really gave us a good sense of security and comfort just knowing that she was being monitored on the other end as well.* [C06, collaborative]

### *Technology Can Act as a Member of the Dyad's Support System*

Medly is a key element of the dyad's support system for 3 reasons: the support provided by Medly is instantaneous; there is a direct point of contact with clinic staff available through the service during business hours; and it facilitates symptom tracking, which reduces the dyad's burden of monitoring:

> *I'm not [asking] every morning, "Well what is it? What are your numbers today?" I might do that for a week or so and then it wears off. I know that if I haven't asked what the numbers are...the coordinator's going to call me and say this has happened and this is what you need to do.* [C13, collaborative]

> *I kind of feel like [Medly]'s almost taking that role of being kind of her caregiver. I'm here for everything else if she ever needed anything, but I almost feel like it's kind of my mom's caregiver to be quite honest. It's kind of taken a bit of that off of me.* [C24, patient-oriented]

Medly's daily input requirements and feedback mechanisms often serve as discussion prompts between the caregiver and care recipient, thus enabling better communication within the dyad:

> *It's also become like there's something in common for both of us to discuss, to analyze, to find out how he feels, is there any difference today and look at the values every day when he takes his reading, are we within the parameters they have given us. And I know that it's not just the value, it also depends on how he feels so there's a conversation always going around that.* [C11, collaborative]

The ICs appreciated Medly for its ability to provide continuity of care while taking over some of the organizational skills that caregiving typically requires, such as providing medication prompts during exacerbation events and medical care in emergencies.

### Technology Can Mitigate the Dyad's Clinical Knowledge Requirements

Empowering patients and formalizing the off-label use of Medly by caregivers can facilitate positive dyadic changes. The knowledge and medical guidance provided by Medly enable a greater understanding of the patient's daily HF status within the dyad:

> *Before being on Medly, he wasn't even sure if he was feeling well or wasn't. So...having some kind of ability to punch in numbers and have an expert look at them and know if they're good or not good was just crucial.* [C14, collaborative]
>
> *On the other side of the app there are human beings who are very well trained and who are specialized in heart failure. I think this program is amazing because of the knowledge base that it has.* [C04, caregiver-oriented]

The constant feedback mechanism associated with Medly's daily check-in prompts facilitates greater awareness of the patient's HF status between care recipients and caregivers:

> *It's really been a very handy tool and it's provided him with the ability to care for himself too. When he's not feeling well, he [uses] this tool as a means of keeping connected with health professionals...He's using it every day, recording his symptoms. It helps him keep in line with, "OK. I need to watch my diet, my fluids, take my medication, check my blood pressure, check my weight," so it's been a very useful tool for him. For me too, because then I know that he's taking check of what's going on with how he's feeling.* [C10, patient-oriented]

Technology can empower ICs to feel more competent and confident in their caregiving abilities by providing a medium to access evidence-based information and actionable feedback for the patient's dynamic HF status. In this way, technology can act as a dyadic self-efficacy tool for HF management.

## Discussion

### Principal Findings

Underlying many of the experiences and anecdotes shared by ICs of patients with HF was the stress of uncertainty regarding a patient's future disease status. HF is a progressive chronic disease that naturally worsens over time. However, progression is nonlinear, and HF exacerbation events cannot always be predicted. This uncertainty is more difficult to cope with if a caregiver within a dyad perceives that they lack the knowledge or support to effectively or adequately care for their care recipient.

Although several IC-facing health technologies exist within the private sector, very few have been formally tested through traditional research. Of the existing technologies that have been evaluated, many have shown to be efficacious in their ability to improve primary and chronic care access, especially in older ethnic adult populations [26]. A systematic review of telehealth interventions built to support ICs showed significant benefits across dyads. The benefits included improved psychological health, confidence, knowledge, patient management, communication with providers, and physical health [28]. Similarly, an experimental study that engaged patients living with HF and family caregivers in an interactive voice response intervention demonstrated the positive effects of dyadic engagement, which resulted in enhanced patient medication adherence, caregiver communication, and quality of life among patients while decreasing the likelihood of HF exacerbations [29].

Therefore, there is an opportunity for technology to better support improved outcomes in HF dyads. For patients experiencing cognitive decline, language barriers, visual impairments or possessing minimal digital literacy, caregivers often stepped in as Medly proxy users on the patient's behalf. In contrast, caregivers in dyads with patients who had few accessibility challenges generally took on a more supportive role and did not directly interact with Medly. Our results suggest that ICs in caregiver-oriented dyads may be more likely to act as proxy users of digital health interventions. With their involvement, patients can overcome this digital divide and adopt digital health interventions. In our study, the ICs across all 3 dyad typologies expressed myriad benefits, from reduced care burden and improved family functioning to an improved sense of security in the care provided to their patient and improved ability to cope with caregiving-related stress. Our findings are aligned with other studies on digital health interventions that formally engage ICs in shared dyadic care and demonstrated improved outcomes; however, this avenue of research remains nascent in application and practice [28].

Our findings further identified a need to focus on improving the dyadic experience as a whole by building IC functionality into digital health self-management interventions. An avenue to address this need would be to formalize the IC role within Medly to address dyadic needs across the spectrum of caregiver engagement (ie, primary or secondary ICs). This approach was supported by the scoping review by Park et al [29] assessing the quality of family-targeted caregiver apps, which highlighted that appraised apps commonly neglect caregiver needs by focusing solely on improving the patient's provision of care.

Although Medly was designed to be patient facing, we have presented scenarios in which Medly was used off label by the IC as a proxy. The dyadic nature of HF caregiving was not initially incorporated into Medly as a design consideration, and we acknowledge that targeted caregiver technology may be more beneficial for certain dyadic typologies. Indeed, feedback from caregivers indicates that the addition of a caregiver account to Medly would be beneficial in diffusing the caregiving burden across the patient's circle of care in times of decline in the health of the patient when the demands of caregiving typically rise. However, the desire for caregiving functionality varied across dyadic typologies. ICs in caregiver-oriented typologies saw less added value in a proxy account as they already used Medly on behalf of the patient through the primary patient account. For them, new caregiver functionality would have limited the utility without additional features that extended beyond the current Medly offering (eg, caregiver mental health support). In contrast, caregivers in collaborative and patient-oriented dyads expressed greater interest in having a caregiver account so that they could remain informed of the care recipient's health status without having to ask the patient directly:

> *It would be great for me to be able to jump on the app and just have a look and see, well, how many pillows did she sleep with last night? Is she having a hard time doing the stairs, that kind of thing...I think she doesn't want to worry me.* [C24, patient-oriented]

Considering the role technology can play in providing peace of mind to dyads, design considerations for accessibility must be prioritized. Inclusive design should be the standard for digital health as it significantly influences the uptake of innovations such as Medly. Failing to proactively consider and design for how older adults and their ICs might use technology for HF management might prevent dyads from meeting their care goals. The ICs in our study informed that their patient partners relied on them to engage with the Medly app because of visual impairments that made it difficult for them to use the app independently. This feedback motivates the need to implement design changes and ensure that the Medly app provides sufficient color contrast, audio commands, font size options, and responsive screen compatibility:

> *My husband is colorblind. So if...there's an alert that comes out in red, for example, he could not see that it's red, and there's a problem. There are programs that you can actually speak and the computer will type it up for you or send a message...I'm wondering if they [could] engage Medly [and it would say], "patient X, what is your weight today? What is your blood pressure today?" and you can verbally respond? And the output would be verbal as well. "Your blood pressure is fine today. Your weight is fine today or your weight is high today, please take two additional ABC medications."* [C09, caregiver-oriented]
>
> *Viewing the results, it's really hard to look at things historically because the screen is so small. It's hard to figure out where he was at certain points because you can't see it on such a small screen.* [C03, caregiver-oriented]

In cases of patient mistrust of the health care system or technology, caregivers may be required to step in and facilitate acceptance of and access to digital health. Such mistrust is especially important to consider in patients belonging to communities that have historically experienced, and continue to experience, challenging and traumatic interactions with these systems, including older adults and those from ethnocultural minority communities. Cultural safety and language accessibility are critical to support the management of chronic conditions that disproportionately affect ethnocultural minority communities [28]. Promoting inclusivity in digital health through equitable design and implementation strategies may simultaneously improve the uptake of remote HF management using technology while mitigating the burden of caregiving on ICs in dyads and enhancing care recipients' autonomy.

## Strengths and Limitations

Although research on IC engagement with DTx is scarce and still emerging, our research team was uniquely positioned and able to explore the experience of ICs with partners enrolled in the Medly program at the UHN. This position allowed us to elucidate the complexity of care needs within the dyadic unit as they sought to comanage HF. These perspectives have not been well represented in prior research, and the portrayal of the experiences of ICs within their HF dyads is a strength of our study. However, this study has several limitations. As a result of our focus on ICs, patient perspectives were not directly assessed as part of this initial phase of our exploratory research. Gathering insights from both entities within the dyadic unit would have provided more holistic insights and is planned for the future phases of this study. In addition, the research team originally sought to explore novel ways of supporting older ethnic adults with limited English proficiency and their ICs in the use of DTx. However, we recognize that most of our IC sample was skewed toward those who self-identified as White (14/20, 70%). Further research on the barriers to accessing digital health within ethnocultural minority populations, which integrates intentional, culturally safe recruitment strategies, is necessary to warrant a more diverse and generalizable sample. In response to this limitation, detailed descriptions of the study results were provided to contextualize the study findings. More generally, future work at the intersection of dyadic HF management and technology use is warranted with other digital health platforms to further evaluate the generalizability of our findings.

## Conclusions

We observed the value of an off-label implementation of the Medly HF DTx in improving the dyadic HF management experience. We believe there is a need to formally and intentionally expand HF technologies to include dyadic needs and goals, considering both the caregiver and the care recipient. When planning technology design and the value of technology to support care dyads in managing chronic conditions, we suggest defining 3 opportunities for support. The first is to identify how technology may be leveraged to increase psychological bandwidth by reducing uncertainty and providing peace of mind. We found that actionable feedback was highly desired by both partners. The second is to develop technology that can serve as a member of the dyad's support system. In our experience, automated prompts for patients to take measurements can mimic the support typically provided by ICs and ease their workload. The third is to consider the ways in which technology can mitigate the dyad's clinical knowledge requirements and learning curves. Our approach was to have real-time actionable feedback paired with a human-in-the-loop, nurse-led model of care. Through a shared model of dyadic care that supports the role of the patient in their own HF management, includes ICs to expand and enhance the patient's capacity to care, and acknowledges the needs of ICs to care for themselves, we anticipate improved outcomes for both partners.

## Acknowledgments

This research was funded by a Canadian Institutes of Health Research Project grant, TRANSFORM HF Collaboration Starter grant, and the AMS Healthcare Fellowship in Compassion and Artificial Intelligence.

## Authors' Contributions

NE, RL, and QP developed the project protocol and contributed to the study design. NE and CB conducted interviews. NE, CB, KY, and KG analyzed the interview transcripts. All authors contributed to manuscript preparation and approved the final version of the manuscript. QP oversaw the research activities and served as the guarantor for this manuscript.

## Conflicts of Interest

NE, KP, CB, KY, RL, AS, and QP are employed by the University Health Network where the Medly system was developed.

Multimedia Appendix 1
Interview guide.
[DOCX File , 24 KB - jmir_v24i9e40108_app1.docx ]

## References

1. Lum J, Hawkin L, Liu J, Ying A, Sladek J, Peckham A. In Focus: Informal caregiving. Ryerson University. 2011 Jul. URL: https://www.ryerson.ca/content/dam/crncc/knowledge/infocus/informalcaregiving/InFocus-InformalCaregiving.pdf [accessed 2022-04-12]
2. Sabia JJ. There's no place like home: a hazard model analysis of aging in place among older homeowners in the PSID. Res Aging 2008 Jan 01;30(1):3-35. [doi: 10.1177/0164027507307919]
3. Ross H, Howlett J, Arnold JM, Liu P, O'Neill B, Brophy J, et al. Treating the right patient at the right time: access to heart failure care. Can J Cardiol 2006 Jul;22(9):749-754. [doi: 10.1016/s0828-282x(06)70290-2]
4. Heart failure. Heart and Stroke Foundation of Canada. URL: https://www.heartandstroke.ca/heart-disease/conditions/heart-failure [accessed 2022-04-13]
5. Heart failure in Canada: complex, incurable and on the rise. Heart and Stroke Foundation of Canada. 2022 Feb 1. URL: https://www.heartandstroke.ca/what-we-do/media-centre/news-releases/heart-failure-in-canada-complex-incurable-and-on-the-rise [accessed 2022-04-12]
6. Buck HG, Mogle J, Riegel B, McMillan S, Bakitas M. Exploring the relationship of patient and informal caregiver characteristics with heart failure self-care using the actor-partner interdependence model: implications for outpatient palliative care. J Palliat Med 2015 Dec;18(12):1026-1032. [doi: 10.1089/jpm.2015.0086] [Medline: 26540092]
7. Buck HG, Stromberg A, Chung ML, Donovan KA, Harkness K, Howard AM, et al. A systematic review of heart failure dyadic self-care interventions focusing on intervention components, contexts, and outcomes. Int J Nurs Stud 2018 Jan;77:232-242. [doi: 10.1016/j.ijnurstu.2017.10.007] [Medline: 29128777]
8. Kitko LA, Hupcey JE, Pinto C, Palese M. Patient and caregiver incongruence in advanced heart failure. Clin Nurs Res 2015 Aug 05;24(4):388-400. [doi: 10.1177/1054773814523777] [Medline: 24599063]
9. Bugajski A, Buck H, Zeffiro V, Morgan H, Szalacha L, Alvaro R, et al. The influence of dyadic congruence and satisfaction with dyadic type on patient self-care in heart failure. Eur J Cardiovasc Nurs 2021 Mar 01;20(3):268-275. [doi: 10.1177/1474515120960002] [Medline: 33611366]
10. Lyons KS, Sadowski T, Lee CS. The role of concealment and relationship quality on patient hospitalizations, care strain and depressive symptoms in heart failure dyads. Eur J Cardiovasc Nurs 2020 Feb 06;19(2):118-124. [doi: 10.1177/1474515119863791] [Medline: 31382771]
11. Molloy GJ, Johnston DW, Witham MD. Family caregiving and congestive heart failure. Review and analysis. Eur J Heart Fail 2005 Jun 28;7(4):592-603. [doi: 10.1016/j.ejheart.2004.07.008] [Medline: 15921800]
12. A roadmap for improving integrated heart failure care in Ontario. CorHealth Ontario. 2019 May. URL: https://www.corhealthontario.ca/Roadmap-for-Improving-Integrated-HF-Care-May-2019.pdf [accessed 2022-04-12]

13. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. User-centered adaptation of an existing heart failure telemonitoring program to ensure sustainability and scalability: qualitative study. JMIR Cardio 2018 Dec 06;2(2):e11466. [doi: 10.2196/11466]
14. Ware P, Ross HJ, Cafazzo JA, Boodoo C, Munnery M, Seto E. Outcomes of a heart failure telemonitoring program implemented as the standard of care in an outpatient heart function clinic: pretest-posttest pragmatic study. J Med Internet Res 2020 Feb 08;22(2):e16538. [doi: 10.2196/16538] [Medline: 32027309]
15. Digital therapeutics definition and core principles. Digital Therapeutics Alliance. 2019 Nov. URL: https://dtxalliance.org/wp-content/uploads/2021/01/DTA_DTx-Definition-and-Core-Principles.pdf [accessed 2022-07-11]
16. Buck HG, Hupcey J, Juárez-Vela R, Vellone E, Riegel B. Heart failure care dyadic typology: initial conceptualization, advances in thinking, and future directions of a clinically relevant classification system. J Cardiovasc Nurs 2019;34(2):159-165. [doi: 10.1097/jcn.0000000000000548]
17. Buck H, Kitko L, Hupcey J. Dyadic heart failure care types: qualitative evidence for a novel typology. J Cardiovasc Nurs 2013;28(6):E37-E46. [doi: 10.1097/JCN.0b013e31827fcc4c] [Medline: 23388704]
18. Van Houtven CH, Voils CI, Weinberger M. An organizing framework for informal caregiver interventions: detailing caregiving activities and caregiver and care recipient outcomes to optimize evaluation efforts. BMC Geriatr 2011 Nov 22;11(1):77. [doi: 10.1186/1471-2318-11-77] [Medline: 22107600]
19. Secules S, McCall C, Mejia JA, Beebe C, Masters AS, L. Sánchez - Peña M, et al. Positionality practices and dimensions of impact on equity research: a collaborative inquiry and call to the community. J Eng Educ 2021 Feb 05;110(1):19-43. [doi: 10.1002/jee.20377]
20. Tracy SJ. Qualitative quality: eight “big-tent” criteria for excellent qualitative research. Qual Inquiry 2010 Oct 01;16(10):837-851. [doi: 10.1177/1077800410383121]
21. Braun V, Clarke V, Hayfield N, Terry G. Thematic analysis. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore: Springer; 2019:843-860.
22. Darwin Holmes AG. Researcher positionality - a consideration of its influence and place in qualitative research - a new researcher guide. Int J Educ 2020 Sep 01;8(4):1-10 [FREE Full text] [doi: 10.34293/education.v8i4.3232]
23. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exercise Health 2021;13(2):201-216. [doi: 10.1080/2159676x.2019.1704846]
24. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) - a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
27. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
28. Pham Q, El-Dassouki N, Lohani R, Jebanesan A, Young K. The future of virtual care for older ethnic adults beyond the COVID-19 pandemic. J Med Internet Res 2022 Jan 07;24(1):e29876. [doi: 10.2196/29876] [Medline: 34994707]
29. Park JY, Tracy CS, Gray CS. Mobile phone apps for family caregivers: a scoping review and qualitative content analysis. Digit Health 2022 Feb 08;8:20552076221076672. [doi: 10.1177/20552076221076672] [Medline: 35154806]

## Abbreviations

**DTx:** digital therapeutic
**HF:** heart failure
**IC:** informal caregiver
**REDCap:** Research Electronic Data Capture
**UHN:** University Health Network

*Edited by G Eysenbach; submitted 08.06.22; peer-reviewed by G Baumblatt, B Riegel, T Phenwan; comments to author 06.07.22; revised version received 27.07.22; accepted 31.07.22; published 07.09.22.*

*Please cite as:*
*El-Dassouki N, Pfisterer K, Benmessaoud C, Young K, Ge K, Lohani R, Saragadam A, Pham Q*
*The Value of Technology to Support Dyadic Caregiving for Individuals Living With Heart Failure: Qualitative Descriptive Study*
*J Med Internet Res 2022;24(9):e40108*
*URL: https://www.jmir.org/2022/9/e40108*
*doi:10.2196/40108*
*PMID:36069782*

Corrigenda and Addenda

# Correction: Ethical Issues in Social Media Recruitment for Clinical Studies: Ethical Analysis and Framework

Bettina M Zimmermann[1,2*], DPhil; Theresa Willem[1,3*], MA; Carl Justus Bredthauer[1], BA; Alena Buyx[1], DrMed

[1]Institute of History and Ethics in Medicine, School of Medicine, Technical University of Munich, Munich, Germany

[2]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

[3]Department of Science, Technology and Society, School of Social Sciences and Technology, Technical University of Munich, Munich, Germany

[*]these authors contributed equally

**Corresponding Author:**
Theresa Willem, MA
Institute of History and Ethics in Medicine
School of Medicine
Technical University of Munich
Prinzregentenstraße 68
Munich, 81675
Germany
Phone: 49 17662315637
Email: theresa.willem@tum.de

**Related Article:**

Correction of: https://www.jmir.org/2022/5/e31231

***(J Med Internet Res 2022;24(9):e40848)*** doi:10.2196/40848

In "Ethical Issues in Social Media Recruitment for Clinical Studies: Ethical Analysis and Framework" (JMIR 2022;24(5):e31231) the authors made one addition to the Acknowledgments section.

The funding note was missing in the originally published article. Therefore, the corrected version of the Acknowledgements' section reads as follows:

> *This work was supported by the European Union's Horizon 2020 research and innovation program (grant agreement no. 848223). The authors thank Dr Nina Goldman for her critical revisions and language editing.*

The correction will appear in the online version of the paper on the JMIR Publications website on September 7, 2022, 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.

## Acknowledgments

This work was supported by the European Union's Horizon 2020 research and innovation program (grant agreement no. 848223). The authors thank Dr Nina Goldman for her critical revisions and language editing.

*Submitted 07.07.22; this is a non–peer-reviewed article;accepted 11.07.22; published 07.09.22.*

*Please cite as:*
*Zimmermann BM, Willem T, Bredthauer CJ, Buyx A*
*Correction: Ethical Issues in Social Media Recruitment for Clinical Studies: Ethical Analysis and Framework*
*J Med Internet Res 2022;24(9):e40848*
*URL: https://www.jmir.org/2022/9/e40848*
*doi:10.2196/40848*
*PMID:36074800*

Corrigenda and Addenda

# Correction: Share to Seek: The Effects of Disease Complexity on Health Information–Seeking Behavior

Ashwag Alasmari[1,2], PhD; Lina Zhou[3], PhD

[1]Computer Science Department, King Khalid University, Abha, Saudi Arabia
[2]University of Maryland, Baltimore County, Baltimore, MD, United States
[3]University of North Carolina at Charlotte, Charlotte, NC, United States

**Corresponding Author:**
Ashwag Alasmari, PhD
University of Maryland, Baltimore County
1000 Hilltop Cir
Baltimore, MD, 21250
United States
Phone: 1 (410) 455 1000
Email: ashwag1@umbc.edu

**Related Article:**

Correction of: https://www.jmir.org/2021/3/e21642/

***(J Med Internet Res 2022;24(9):e42949)*** doi:10.2196/42949

In "Share to Seek: The Effects of Disease Complexity on Health Information–Seeking Behavior" (J Med Internet Res 2021;23(3):e21642), the authors noted two errors.

In the originally published article, affiliations of the first author, Ashwag Alasmari, appeared in the following order:

> *Ashwag Alasmari[1,2]*
>
> *[1]University of Maryland, Baltimore County, Baltimore, MD, United States*
>
> *[2]King Khalid University, Abha, Saudi Arabia*

Moreover, the original Affiliation 2 inadvertently did not specify any department.

In the corrected version, affiliations appear in the following order, and the reordered Affiliation 1 specifies the author's affiliated department:

> *Ashwag Alasmari[1,2]*
>
> *[1]Computer Science Department, King Khalid University, Abha, Saudi Arabia*
>
> *[2]University of Maryland, Baltimore County, Baltimore, MD, United States*

The correction will appear in the online version of the paper on the JMIR Publications website on September 30, 2022, 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 25.09.22; this is a non–peer-reviewed article;accepted 27.09.22; published 30.09.22.*

*Please cite as:*
*Alasmari A, Zhou L*
*Correction: Share to Seek: The Effects of Disease Complexity on Health Information–Seeking Behavior*
*J Med Internet Res 2022;24(9):e42949*
*URL: https://www.jmir.org/2022/9/e42949*
*doi:10.2196/42949*
*PMID:36179345*

Original Paper

# The Impact of the COVID-19 Pandemic on Internet Use and the Use of Digital Health Tools: Secondary Analysis of the 2020 Health Information National Trends Survey

Billy Zeng[1], BA, MD; Natalie A Rivadeneira[2,3], MPH; Anita Wen[4], BS; Urmimala Sarkar[2,3], MPH, MD; Elaine C Khoong[2,3], MS, MD

[1]School of Medicine, University of California San Francisco, San Francisco, CA, United States

[2]Division of General Internal Medicine at Zuckerberg San Francisco General Hospital, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[3]Center for Vulnerable Populations at Zuckerberg San Francisco General Hospital, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[4]Department of Nutrition, University of California Davis, Davis, CA, United States

**Corresponding Author:**
Elaine C Khoong, MS, MD
Division of General Internal Medicine at Zuckerberg San Francisco General Hospital
Department of Medicine
University of California San Francisco
1001 Potrero Avenue
Bldg 10, Ward 13
San Francisco, CA, 94110
United States
Phone: 1 6282063188
Email: elaine.khoong@ucsf.edu

## Abstract

**Background:** The COVID-19 pandemic increased the use of digital tools in health care (eg, patient portal, telemedicine, and web-based scheduling). Studies have shown that older individuals, racial/ethnic minority groups, or populations with lower educational attainment or income have lower rates of using digital health tools. Digitalization of health care may exacerbate already existing access barriers in these populations.

**Objective:** This study evaluated how use of digital tools to asynchronously communicate with clinicians, schedule appointments, and view medical records changed near the beginning of the pandemic.

**Methods:** Using 2020 Health Information National Trends Survey (HINTS) data, we examined internet use and 7 digital health technology use outcomes (electronic communication with a provider, electronic appointment scheduling, electronic test result viewing, patient portal access, portal use to download health records, portal use for patient-provider communication, and portal use to view test results). The HINTS surveyors designated surveys received after March 11, 2020, as postpandemic responses. Using weighted logistic regression, we investigated the impact of the pandemic after adjusting for sociodemographic traits (age, race/ethnicity, income, education, and gender), digital access (having ever used the internet and smartphone/tablet ownership), and health-related factors (insurance coverage, caregiver status, having a regular provider, and chronic diseases). To explore differences in changes in outcomes among key sociodemographic groups, we tested for significant interaction terms between the pandemic variable and race/ethnicity, age, income, and educational attainment.

**Results:** There were 3865 respondents (1437 prepandemic and 2428 postpandemic). Of the 8 outcomes investigated, the pandemic was only significantly associated with higher odds (adjusted odds ratio 1.99, 95% CI 1.18-3.35) of using electronic communication with a provider. There were significant interactions between the pandemic variable and 2 key sociodemographic traits. Relative to the lowest income group (<US $20,000), the highest income group (≥US $75,000) had increased growth in the odds of ever having used the internet in postpandemic responses. Compared to the most educated group (postbaccalaureates), groups with lower educational attainment (high school graduates and bachelor’s degree) had lower growth in the odds of using electronic communication with a provider in postpandemic responses. However, individuals with less than a high school degree had similar growth to the postbaccalaureate group in using electronic communication with a provider.

**Conclusions:** Our study did not show a widespread increase in use of digital health tools or increase in disparities in using these tools among less advantaged populations in the early months of the COVID-19 pandemic. Although some advantaged populations reported a greater increase in using the internet or electronic communication with a provider, there were signs that some less advantaged populations also adapted to an increasingly digital health care ecosystem. Future studies are needed to see if these differences remain beyond the initial months of the pandemic.

***(J Med Internet Res 2022;24(9):e35828)*** doi:10.2196/35828

**KEYWORDS**

COVID-19; digital divide; eHealth; telehealth

## Introduction

In response to the COVID-19 public health emergency, many American health centers transitioned to telemedicine almost overnight, with most visits conducted over the phone or video and only a limited number of visits were conducted in person [1]. As more health care became web-based, digital health tools became even more important [2]. However, studies have shown that individuals who are older, identify as persons of color, or have lower educational attainment or income are less likely to use a variety of digital health tools (eg, mobile health apps, telemedicine, and web-based medical records) [3-5].

Many of these differences in the use of digital health tools stem from structural factors—including the cost of internet access, broadband infrastructure, and digital literacy skills [6,7]—that were largely unaddressed during the pandemic. Advocates for health care equity therefore worried that requiring health care services to be accessed through digital tools without interventions to address structural barriers to digital equity in diverse populations would exacerbate preexisting differences in the use of digital health tools [8].

Although much of the focus on digital health equity since the start of the COVID-19 pandemic has been on web-based visits (or telemedicine) and increasingly remote patient monitoring tools, digital tools support a variety of other health-related tasks. Digital health technologies have been defined to include “mobile health (mHealth), health information technology, wearable devices, telehealth and telemedicine and personalized medicine” [8]. Patients can perform many health care–related tasks using digital health technologies; for example, patients can use digital tools to asynchronously communicate with clinicians, view their web-based medical records or test results, or schedule appointments.

The Health Information National Trends Survey (HINTS) is a nationally administered annual survey from the National Cancer Institute, which collects information about health communication, including patients’ use of technology for health care–related tasks outside of web-based visits [9,10]. Beginning in 2008, HINTS began including questions related to the use of the internet and other digital tools for supporting health care–related tasks. Therefore, the HINTS survey data provided an opportunity to gain insights into whether digital equity concerns were created in the early stages of the pandemic for non–telemedicine-related digital health tasks (eg, messaging with a clinician, scheduling appointments, and viewing web-based medical records).

Using HINTS data, we investigated whether disparities increased in the use of digital tools to conduct health care–related tasks after the start of the COVID-19 public health emergency. (Within this paper, we will use the term “disparity” to describe differences between groups.)

We focused on 4 sociodemographic factors previously documented to be associated with disparities in using digital health tools: age, race/ethnicity, education, and income [3-5,11-13]. We specifically hypothesized that in the early stages of the pandemic during which HINTS 2020 data were collected, there may have been increased disparities in the use of the internet and digital tools for health care–related tasks, with lower use in populations who are older, are in racial/ethnic minority groups, have lower educational attainment, or have lower income.

## Methods

### Survey Administration

Details about the HINTS administration and design are publicly available [10]. In brief, English and Spanish surveys are sent out randomly to US residential addresses and returned via mail or the internet. HINTS collects information on internet use and the use of digital tools to conduct health care tasks such as communicating with doctors, making appointments, and viewing test results [10]. We used HINTS 5 cycle 4 data, which were collected between February and June 2020 [14]. These surveys were sent to a random sample of addresses with an oversampling of areas with high minority populations to increase precision for inference on minority populations. Survey sample weights are provided in the data to allow for inferences about the whole US population. The 2020 HINTS included a variable to indicate if the survey was returned before or after the COVID-19 pandemic; surveys received after March 11, 2020, were flagged by the HINTS surveyors as postpandemic. The survey response rate for HINTS in 2020 was 37%, which was consistent with prior years.

### Outcomes

We selected 8 dichotomous (yes/no) outcome variables from questions about having ever used the internet and the use of digital tools for health-related tasks (see Multimedia Appendix 1). The internet use outcome asked if respondents had ever used the internet. There were 3 outcome variables that focused on the use of electronic means to talk to a doctor, make an appointment with a health care clinician, or view test results within the past 12 months. The last 4 outcome variables pertained to patient portal use: accessing their web-based patient portal, downloading health records, communicating with a provider, or viewing test results within the past 12 months. The last 3 outcome variables on the use of patient portals for various tasks were only asked of respondents who reported having accessed their patient portal.

### Model Design

To guide our analysis, we conceptualized the predictors that could impact each of these outcomes. In addition to having the pandemic as a key predictor variable in all models, we identified 3 groups of predictors (sociodemographic traits, digital access, and health-related factors) drawn from prior literature and described below [15-18]. Of note, the having ever used the internet outcome was included as 1 of the digital access predictor variables for modeling the other 7 outcomes.

### Predictor Variables and Covariates

#### *The COVID-19 Pandemic*

The pandemic was a key predictor variable that indicated if the survey response occurred after (survey received after March 11, 2020) or before the COVID-19 pandemic. This designation was made by the HINTS surveyors.

#### *Sociodemographic Traits*

The sociodemographic traits included in the model were age (18-34, 35-49, 50-64, 65-74, and ≥75 years), race/ethnicity (Asian, Black, Hispanic of any race, non-Hispanic White, and other), education (less than a high school degree, high school graduate, some college, bachelor’s degree, and postbaccalaureate), income (<US $20,000, US $20,000-$34,999, US $35,000-$49,999, US $50,000-$74,999, and >US $75,000), and gender (male and female). All predictors were categorical variables. Missing values in the income data were imputed and supplied by the HINTS data set. For the logistic regression models, the reference groups for age, race/ethnicity, education, income, and gender were the following, respectively: aged 18-34 years, non-Hispanic White, postbaccalaureate education, income<US $20,000, and male.

#### *Digital Factors*

There were 2 dichotomous variables included in this group: owns a tablet or smartphone and having ever used the internet. Having ever used the internet was an outcome in 1 model but was included as a covariate in the other models.

#### *Health Care Factors*

There were 3 dichotomous health care–related variables: functions as a caregiver for another individual, has access to a regular provider, and has insurance. We also included 1 categorical variable: the number of chronic diseases (0, 1, 2, or ≥3) based on self-reported diagnoses of depression, hypertension, diabetes, heart disease, or lung disease, with 0 chronic diseases used as the reference value.

### Analysis

We report descriptive statistics of predictor variables, covariates, and outcomes unweighted. To infer population-level statistics, we report weighted proportions using weights provided by the HINTS data set. Using weight adjusted survey data, we constructed bivariate and multivariable logistic regression models for each of the 8 outcomes. The models for all outcomes used all the predictor and covariate variables listed above; we did not conduct variable selection, since all variables have been shown to impact these outcomes in the literature.

To determine the impact of the pandemic, we focused on the pandemic variable and the interaction terms between the pandemic variable and the 4 sociodemographic traits of interest (race/ethnicity, age, education, and income). The Wald test was used to evaluate the interaction between pandemic status and these 4 sociodemographic traits. Interactions at $P<.10$ were included in the final overall model. The final overall model was used to generate marginal expected odds and SE of each outcome for each sociodemographic and pandemic interaction pair at $P<.10$.

All analyses were performed using R statistical software (version 4.1.0; R Foundation for Statistical Computing). To adjust for complex survey design, we used a survey adjustment via the *survey* package (version 4.0) to apply sampling and jackknife replicate weights [19]. The *survey* package was also used to conduct the statistical Wald test for interaction variables. All regression models were created with complete cases. The *emmeans* package (version 1.6.0) was used to compute the expected means and SEs of odds [20], and the *ggplot2* package (version 3.3.3) was used for plot generation [21].

We used $P<.05$ to determine statistical significance for all outcomes. We did not adjust for multiple hypothesis testing due to having planned few comparisons rather than every possible comparison and to avoid increasing type II error [22,23].

## Results

### Survey Respondents and Outcomes

Of the 3865 survey respondents, 1437 responded before the pandemic indicator and 2428 responded post the pandemic. Table 1 describes the survey participants and outcomes. A large portion (3148/3865, 86%) of participants reported having ever used the internet, but less than half of them reported using any of the digital health tools. The most common uses of digital tools for health care–related tasks were using electronic means to schedule a health care appointment (1891/3865, 49%) and communicate with a provider (1800/3865, 47%). Only 39% (1553/3865) of respondents reported having ever accessed their patient portal. (As noted in table 1, percentages are weighted so may not align with the n/N presented.) Of those who accessed their patient portal, 87% (1349/1553) reported viewing their test results, and 59% (920/1553) reported messaging their clinicians.

**Table 1.** Traits of included participants (N=3865).

| Trait, variable[a] | 2020, prepandemic (n=1437), n (weighted %[b]) | 2020, postpandemic (n=2428), n (weighted %[b]) | 2020, total, n (weighted %[b]) |
|---|---|---|---|
| **Sociodemographic** | | | |
| **Gender** | | | |
| Female | 804 (47.71) | 1400 (51.59) | 2204 (50.22) |
| **Age (years)** | | | |
| 18-34 | 151 (19.21) | 333 (28.89) | 484 (25.47) |
| 35-49 | 212 (21.93) | 491 (26.37) | 703 (24.80) |
| 50-64 | 433 (31.88) | 709 (24.25) | 1142 (26.95) |
| 65-74 | 361 (13.81) | 508 (10.42) | 869 (11.62) |
| ≥75 | 237 (9.74) | 303 (7.61) | 540 (8.36) |
| **Race/ethnicity** | | | |
| Asian | 51 (3.84) | 110 (5.37) | 161 (4.83) |
| Black | 135 (7.70) | 346 (11.75) | 481 (10.32) |
| Hispanic | 170 (11.86) | 426 (17.84) | 596 (15.73) |
| White | 904 (7.37) | 1229 (7.31) | 2133 (7.34) |
| Other | 49 (4.25) | 70 (2.45) | 119 (3.09) |
| **Income (US $)** | | | |
| <20,000 | 258 (15.13) | 506 (17.38) | 764 (16.58) |
| 20,000-34,999 | 189 (11.02) | 302 (11.73) | 491 (11.48) |
| 35,000-49,999 | 180 (11.74) | 336 (12.74) | 516 (12.39) |
| 50,000-74,999 | 257 (17.44) | 392 (17.98) | 649 (17.79) |
| ≥75,000 | 547 (43.66) | 880 (39.67) | 1427 (41.08) |
| **Education** | | | |
| Less than a high school degree | 90 (7.01) | 183 (8.25) | 273 (7.81) |
| High school graduate | 251 (19.69) | 454 (23.09) | 705 (21.89) |
| Some college | 415 (38.61) | 666 (37.82) | 1081 (38.10) |
| Bachelor’s degree | 358 (19.97) | 621 (17.32) | 979 (18.26) |
| Postbaccalaureate | 285 (12.05) | 399 (10.70) | 684 (11.18) |
| **Digital Factors** | | | |
| Having ever used the internet | 1187 (87.09) | 1961 (85.09) | 3148 (85.80) |
| Owns a tablet or smartphone | 1210 (87.63) | 2029 (88.58) | 3239 (88.25) |
| **Health Factors** | | | |
| Has insurance | 1352 (90.43) | 2252 (89.42) | 3604 (89.78) |
| Has a regular provider | 1046 (69.60) | 1582 (56.91) | 2628 (61.39) |
| Is a caregiver | 198 (14.53) | 378 (16.66) | 576 (15.91) |
| **Number of Chronic Disease** | | | |
| 0 | 506 (42.58) | 922 (45.45) | 1428 (44.44) |
| 1 | 356 (20.76) | 550 (18.98) | 906 (19.61) |
| 2 | 335 (20.58) | 560 (21.88) | 895 (21.42) |
| ≥3 | 229 (15.71) | 362 (12.79) | 591 (13.82) |
| **Outcomes** | | | |
| Electronic communication with a provider | 659 (48.22) | 1141 (45.61) | 1800 (46.53) |

| Trait, variable[a] | 2020, prepandemic (n=1437), n (weighted %[b]) | 2020, postpandemic (n=2428), n (weighted %[b]) | 2020, total, n (weighted %[b]) |
|---|---|---|---|
| Electronic means to make Appointments | 680 (48.78) | 1211 (48.73) | 1891 (48.75) |
| Electronic means to view test results | 634 (45.67) | 995 (39.42) | 1629 (41.63) |
| Having ever accessed their patient portal | 605 (41.18) | 948 (38.57) | 1553 (39.49) |
| Patient portal to message a provider[c] | 350 (57.85) | 570 (60.13) | 920 (59.24) |
| Patient portal to view test results[c] | 530 (87.60) | 819 (86.39) | 1349 (86.53) |
| Patient portal to download health records[c] | 171 (28.26) | 284 (29.95) | 455 (29.30) |

[a] Each variable had less than 10% missing data.

[b] The percentage rates were calculated using weighted data to represent the US population.

[c] Patient portal tasks were only asked of those who had accessed the patient portal. Therefore, the proportions are reported only out of those that reported having ever accessed their patient portal.

## Impact of the Pandemic

All bivariate models and multivariable analysis are shown in Multimedia Appendices 2 and 3, respectively. In the adjusted analysis, older age, lower income, lower educational attainment, and race/ethnic minority groups were associated with lower odds of having ever used the internet (Multimedia Appendix 3). The same patterns for age, income, and educational attainment were seen for the other outcomes, but the findings were mixed by race/ethnicity (Multimedia Appendix 3).

After accounting for other variables, the pandemic variable was only significant for using electronic means to communicate with a provider. Postpandemic respondents had higher odds (adjusted odds ratio [aOR] 1.99, 95% CI 1.18-3.35; *P*=.01; see Multimedia Appendix 3) of using electronic means to communicate with a provider than responses from the prepandemic period.

The interaction between the pandemic variable and 4 sociodemographic variables (age, race/ethnicity, education, and income) was only significant for 2 outcomes: having ever used the internet and electronic communication with a provider (see Multimedia Appendix 3.). These significant interactions are shown in Figures 1 and 2.

For the outcome related to internet use, there was a significant interaction between the pandemic variable and income (see Figure 1). Specifically, respondents in the ≥US $75,000 income group had an increase in the odds of having ever used the internet post the pandemic, which was significantly different from the <US $20,000 income group (*P*=.02).

The use of electronic communication with a provider was notably increased in the highest educational attainment group (postbaccalaureate). As seen in Figure 2 (and detailed in Multimedia Appendix 3), the highest educated group (postbaccalaureate) had a significantly greater growth in odds of conducting this digital task than both the groups with a bachelor's degree (*P*=.02) and high school education only (*P*=.01). There was also a trend toward greater growth compared to the group with some college education (aOR 0.54, 95% CI 0.27-1.06), but this result was not statistically significant (*P*=.07). The growth in using electronic communication with a provider in the most educated group was not significantly different from patterns seen in conducting this task by the respondents who reported less than a high school education (*P*=.80).

**Figure 1.** Odds of having ever used the internet before (pre) and after (post) the pandemic among income groups (in US $).

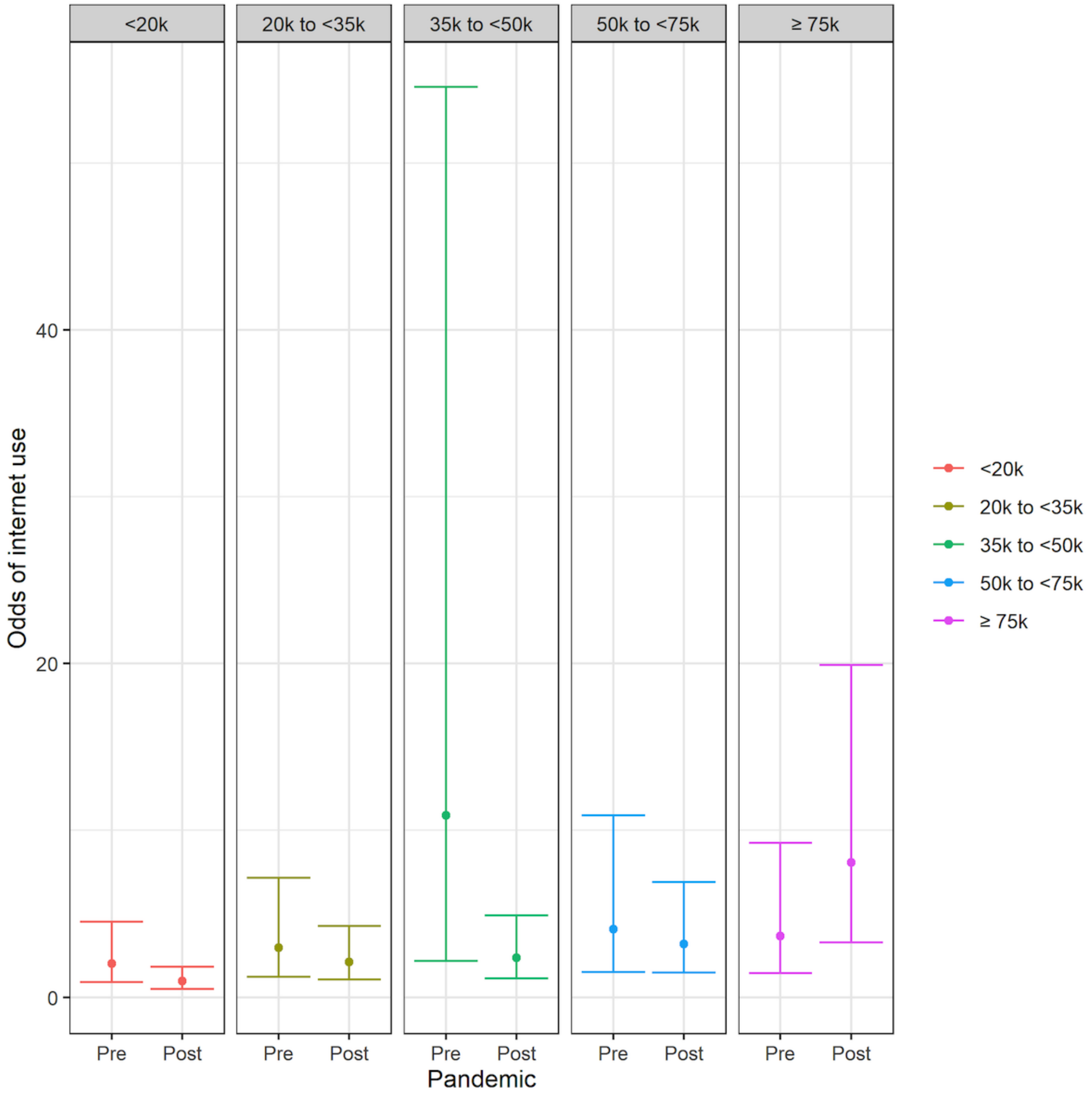
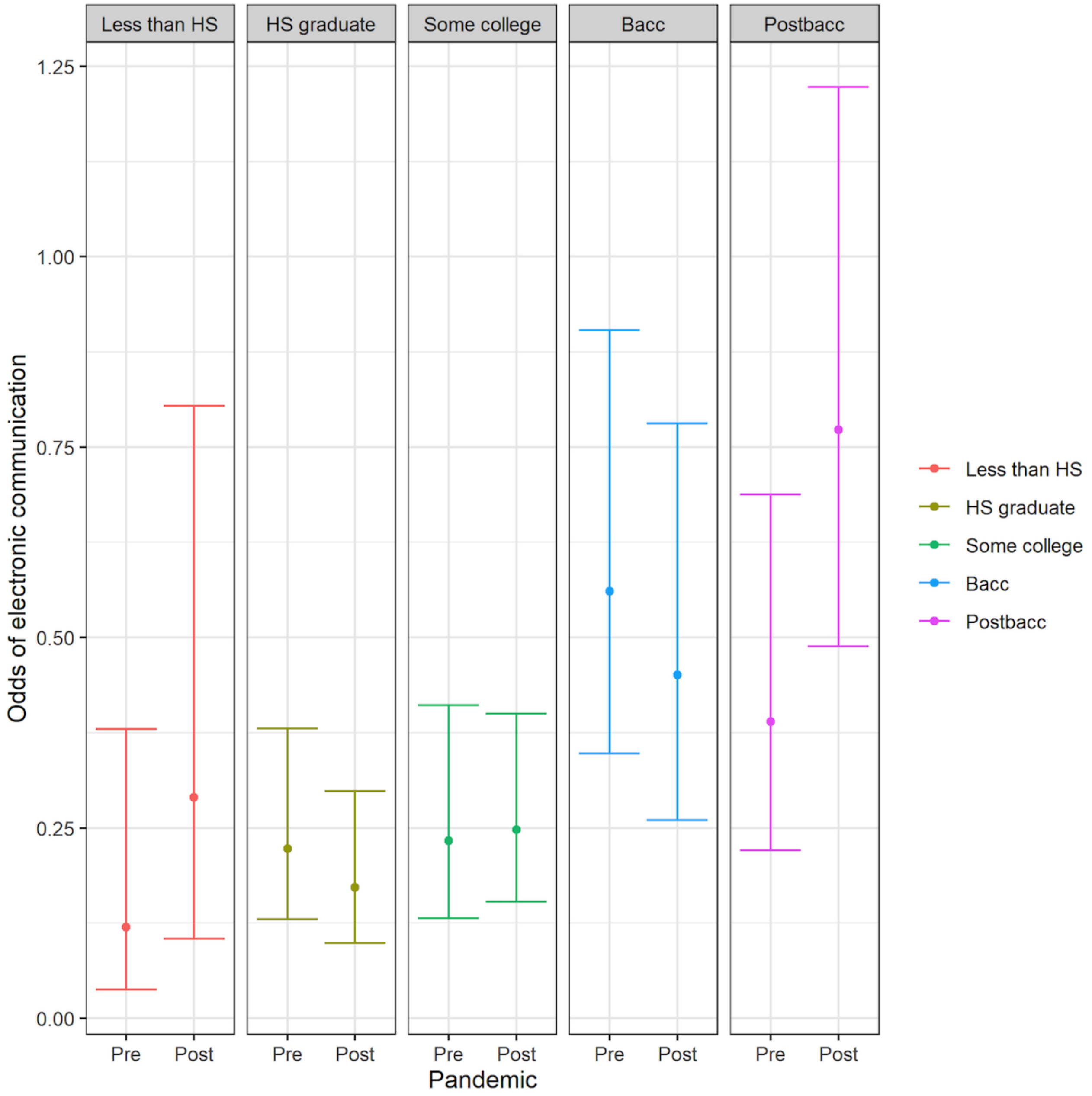


**Figure 2.** Odds of using electronic communication with a provider among different education groups before (pre) and after (post) the pandemic. Bacc: baccalaureate; HS: high school; Postbacc: postbaccalaureate.



## Discussion

### Principal Findings

Overall, we found mixed results on how the pandemic affected internet use, the use of digital tools to communicate with clinicians or schedule appointments, and patient portal use. For most of the outcomes, there were no significant differences before and after the pandemic in the early months of the pandemic and no significant changes in disparities in the uptake of digital health tools.

Consistent with prior literature, we did find that populations with a history of digital exclusion (older, lower income, lower educational attainment, and racial/ethnic minority groups) continue to have lower odds of using the internet and a variety of digital health tools. These disparities, particularly in telemedicine use, have been repeatedly documented since the start of the pandemic [24-29]. However, unlike most other studies, we did not study telemedicine use. In prior studies, we have found that the accessibility and use of 1 digital health tool does not translate to other tools [4]; that is, the lessons learned about disparities in telemedicine use may not be applicable to disparities in patient portal use or the use of web-based communication and scheduling tools. By focusing on web-based scheduling, electronic communication, and patient portal use—tools that were relatively widely available both before and after the pandemic, we were able to explore how the pandemic immediately changed the use of these tools and if there was an increase or decrease in disparities in using these digital tools. This comparison of disparities before and after the pandemic on nontelemedicine digital health technologies is different from much of the literature that has largely provided only a static look at the disparities in telemedicine uptake post the pandemic.

With this focus in mind, we did find that immediately after the pandemic, after adjusting for other factors, there were increased odds overall in the use of electronic communication with a provider. One reason for this finding may be that the policies enacted by the Centers for Medicare and Medicaid Services to incentivize the use of telehealth [30,31] created a digital health care environment that made electronic communication more accessible to all populations. Alternatively, it may also suggest that the move to telehealth made it more necessary or important for all individuals, including those who had not previously used electronic communication with their clinicians, to use these telecommunication tools to seek health care or advice from a medical professional.

Our study had mixed findings on how differences in the uptake of these digital tools were immediately impacted by the public health emergency. Immediately after the start of the pandemic, the highest income group (≥US $75,000) had a greater rate of growth in having ever used the internet than the lowest income group (<US $20,000), suggesting a widening of the disparity between income groups. This finding may reflect that higher income earners were more likely to have jobs that could be performed remotely through the internet than lower income groups [32]. A 2019 survey from Pew Research showed that 98% of respondents with an income >US $75,000 used the internet in contrast to 82% of respondents with an income <US $30,000; given the already high rate of internet use in high-income households, it seemed that there should be little room for additional growth among high-income earners. This finding reinforces the need to ensure that structural barriers to accessing the internet for low-income households are mitigated [7].

In contrast to the findings among income groups, there was some suggestion of the gaps closing between groups with different levels of educational attainment. Both the lowest educational attainment respondents (less than high school) and highest educational attainment respondents (postbaccalaureate) had similar rates of growth in the use of electronic communication tools (eg, smartphones, internet, and email) with their doctors. However, the bachelor's degree holders and high school graduates had decreases in the odds of using electronic communication with their doctors after the pandemic, which were significantly different from the most educated group. Together, these findings suggest that although some disparities in the use of electronic communication with clinicians were closing, others were widening. It is worth highlighting that the most vulnerable group from an educational attainment perspective (less than high school education) had a larger growth in using electronic communication tools with their clinician relative to most other respondents, which defies a frequent pattern of innovations disseminating the most slowly to the most disadvantaged.

Given the rapid move of health care to telehealth settings [1], we hypothesized there would be increased inequities after the pandemic in most of our outcomes related to the use of digital tools for health care tasks. However, we had few significant findings except for respondents who reported having ever used the internet or using electronic communication with a provider. Since these data are from early in the pandemic, these 2 digital tasks likely serve as the earliest indicators of how populations were adapting to an increasingly digital health care ecosystem. We anticipate that as more data become available, we may see more changes in the use of the other digital health technologies, including those evaluated in this study as well as other tools such as telemedicine or remote patient monitoring. Although we are somewhat reassured that in the early days of the pandemic, these data do not suggest a consistent widening of inequities between more advantaged and less advantaged populations [2,6,8,33], we also believe it is necessary to reevaluate these outcomes later in the pandemic once health care teams and patients had become more accustomed to conducting more health care tasks remotely. Researchers have documented that the use of telemedicine immediately at the beginning of the COVID-19 pandemic did not reflect the more long-term patterns on telemedicine use [34]. We suspect that similar patterns for other digital health technologies may also emerge.

We believe it important to specifically highlight that we found no changes in any of the patient portal tasks, despite patient portals being the primary digital health tool that has been adopted by health systems to increase patient engagement and care accessibility. Many health care systems already had patient portals in place and tried to use their patient portals to address health care needs during the pandemic; however, studies have repeatedly showed the significant barriers to using a patient portal, including the lack of technical skills, usability, privacy concerns, and the lack of physician encouragement [35,36]. This study suggests that even in an environment where the use of a patient portal may be even more important, patient portal products did not address patients' needs early in the pandemic; this finding is reinforced by the multiple health care systems that found that when they used patient portals to address COVID-19–related care needs (testing and vaccine scheduling), there was inequitable access to care [37]. Since the start of the pandemic, patient portal products have attempted to become more patient-centered [38], and health care teams have increased efforts to improve access for historically excluded populations [39]; future studies should evaluate if these efforts have had the intended impact of reducing disparities in patient portal use.

### Limitations

This study has several limitations. Since the 2020 HINTS responses were collected in a 5-month period between February and June 2020, the results only reflect the early impact of the pandemic. In addition, most outcome questions inquired about electronic communication over the last 12 months, hence outcomes may be less sensitive to the immediate behavior changes resulting from the pandemic. For patient portal–related outcomes, the sample size was limited to respondents who had accessed their patient portal; therefore, there may have been inadequate power to detect statistically significant changes in patient portal use. Although the survey weights are designed to extrapolate these data to the American population, owing to the limited sample size in some subgroups, there may not be enough variability to accurately evaluate the outcomes. For example, all Asian individuals in the postpandemic group reported the use of the patient portal for viewing a test result (Multimedia Appendix 3), suggesting an inadequate diversity of HINTS respondents, and these results should be considered with caution [40]. Furthermore, some studies have shown that postpandemic survey respondents are different from prepandemic survey respondents, and therefore, caution must be used when comparing responses in this survey to those in prior HINTS cycles [41]. However, we are reassured that the response rate for this HINTS cycle was similar to prior years [14]. Despite these limitations, this study adds value to the literature by evaluating early changes in the use of digital health tools during the pandemic and focusing explicitly on changes in use among historically excluded populations.

## Conclusions

Our study finds that early within the pandemic, there was not widespread increase in the use of digital health tools or in disparities in the use of digital health tools. Although these data were only from the first 3 months of the pandemic, we did find an increase in odds of using electronic communication with a provider after the pandemic and some mixed results on whether preexisting inequities between groups in the use of digital health increased. Despite health care systems’ reliance on patient portals to increase patient access and engagement, we did not see changes in the use of patient portals during the early stages of the pandemic. These early data from the pandemic support the need to explicitly study a wide range of digital health care–related tasks. Changes in the use of 1 digital task may not translate to other health care–related digital tasks.

## Acknowledgments

We thank Charles McCulloch for his advice on this study. Research reported in this publication was supported by the National Heart Lung and Blood Institute of the National Institutes of Health (NIH) under award K12HL138046 and K23HL157750 (ECK); the National Center for Advancing Translational Sciences of the NIH under award KL2TR001870 (ECK); and the National Cancer of the NIH under award K24CA212294 (US). The content is solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

## Authors' Contributions

BZ contributed to the conception and design of the study, analysis of the data, and drafting the work. NAR contributed to the design of the study, interpretation of the data, and critically revising the work. AW contributed to the analysis of the data and drafting of the work. US contributed to the interpretation of the data and critically revising the work. ECK contributed to the conception and design of the study, interpretation of the data, and critically revising the work. All authors approved the final version of the manuscript.

## Conflicts of Interest

US has received funding from AppliedVR, InquisitHealth, RecoverX, and Somnology (research contracts). She has also received funding from The Doctors Company (gift), the American Medical Association's Equity and Innovation Advisory Group (honoraria), and Hopelab (grant).

Multimedia Appendix 1
Survey questions and outcome variables.
[DOCX File , 14 KB - jmir_v24i9e35828_app1.docx ]

Multimedia Appendix 2
Unadjusted bivariate models.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v24i9e35828_app2.xlsx ]

Multimedia Appendix 3
Multivariable models.
[XLSX File (Microsoft Excel File), 18 KB - jmir_v24i9e35828_app3.xlsx ]

## References

1. Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. BMJ Health Care Inform 2020 Jun 21;27(1):e100166 [FREE Full text] [doi: 10.1136/bmjhci-2020-100166] [Medline: 32565418]
2. Lyles CR, Wachter RM, Sarkar U. Focusing on digital health equity. JAMA 2021 Dec 09;326(18):1795-1796. [doi: 10.1001/jama.2021.18459] [Medline: 34677577]
3. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]

4. Khoong EC, Rivadeneira NA, Hiatt RA, Sarkar U. The use of technology for communicating with clinicians or seeking health information in a multilingual urban cohort: cross-sectional survey. J Med Internet Res 2020 Apr 06;22(4):e16951 [FREE Full text] [doi: 10.2196/16951] [Medline: 32250280]
5. Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]
6. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021 Mar 17;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
7. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st Century Cures Act era. JAMA 2020 Jul 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
8. Lee P, Abernethy A, Shaywitz D, Gundlapalli A, Weinstein J, Doraiswamy PM, et al. Digital health COVID-19 impact assessment: lessons learned and compelling needs. NAM Perspect 2022 Jan 18;2022 [FREE Full text] [doi: 10.31478/202201c] [Medline: 35402858]
9. 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]
10. About HINTS. National Cancer Institute. URL: https://hints.cancer.gov/about-hints/learn-more-about-hints.aspx [accessed 2021-05-04]
11. Davies AR, Honeyman M, Gann B. Addressing the digital inverse care law in the time of COVID-19: potential for digital technology to exacerbate or mitigate health inequalities. J Med Internet Res 2021 Apr 07;23(4):e21726 [FREE Full text] [doi: 10.2196/21726] [Medline: 33735096]
12. 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 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
13. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50. [doi: 10.2196/jmir.5105] [Medline: 26944212]
14. Health Information National Trends Survey 5 (HINTS 5) Cycle 4 Methodology Report. National Cancer Institute. 2020 Dec. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS5_Cycle4_MethodologyReport.pdf [accessed 2022-08-31]
15. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jul 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
16. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 01;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
17. Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, et al. Growth dynamics of patient-provider internet communication: trend analysis using the Health Information National Trends Survey (2003 to 2013). J Med Internet Res 2018 Mar 29;20(3):e109 [FREE Full text] [doi: 10.2196/jmir.7851] [Medline: 29599107]
18. Bangerter LR, Griffin J, Harden K, Rutten LJ. Health information-seeking behaviors of family caregivers: analysis of the Health Information National Trends Survey. JMIR Aging 2019 Jan 14;2(1):e11237 [FREE Full text] [doi: 10.2196/11237] [Medline: 31518309]
19. Lumley T. survey: analysis of complex survey samples. The Comprehensive R Archive Network. 2021 Jul 19. URL: https://CRAN.R-project.org/package=survey [accessed 2022-08-31]
20. Lenth RV, Buerkner P, Herve M, Jung M, Love J, Miguez F, et al. emmeans: estimated marginal means, aka least-squares means. The Comprehensive R Archive Network. 2021. URL: https://CRAN.R-project.org/package=emmeans [accessed 2021-05-26]
21. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, RStudio. ggplot2: create elegant data visualisations using the grammar of graphics. The Comprehensive R Archive Network. 2020. URL: https://CRAN.R-project.org/package=ggplot2 [accessed 2021-05-26]
22. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990 Jan;1(1):43-46. [Medline: 2081237]
23. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998 May 18;316(7139):1236-1238 [FREE Full text] [doi: 10.1136/bmj.316.7139.1236] [Medline: 9553006]
24. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the COVID-19 pandemic. NEJM Catal Innov Care Deliv 2020 May 04 [FREE Full text]
25. Khoong EC, Butler BA, Mesina O, Su G, DeFries TB, Nijagal M, et al. Patient interest in and barriers to telemedicine video visits in a multilingual urban safety-net system. J Am Med Inform Assoc 2021 Feb 15;28(2):349-353 [FREE Full text] [doi: 10.1093/jamia/ocaa234] [Medline: 33164063]
26. Rodriguez JA, Betancourt JR, Sequist TD, Ganguli I. Differences in the use of telephone and video telemedicine visits during the COVID-19 pandemic. Am J Manag Care 2021 Jan;27(1):21-26 [FREE Full text] [doi: 10.37765/ajmc.2021.88573] [Medline: 33471458]

27. Chunara R, Zhao Y, Chen J, Lawrence K, Testa PA, Nov O, et al. Telemedicine and healthcare disparities: a cohort study in a large healthcare system in New York City during COVID-19. J Am Med Inform Assoc 2021 Jan 15;28(1):33-41 [FREE Full text] [doi: 10.1093/jamia/ocaa217] [Medline: 32866264]
28. Ruberto RA, Schweppe EA, Ahmed R, Swindell HW, Cordero CA, Lanham NS, et al. Disparities in telemedicine utilization during COVID-19 pandemic: analysis of demographic data from a large academic orthopaedic practice. JB JS Open Access 2022 Apr 08;7(2):e21.00116 [FREE Full text] [doi: 10.2106/JBJS.OA.21.00116] [Medline: 35425872]
29. Chang JE, Lai AY, Gupta A, Nguyen AM, Berry CA, Shelley DR. Rapid transition to telehealth and the digital divide: implications for primary care access and equity in a post-COVID era. Milbank Q 2021 Jun;99(2):340-368 [FREE Full text] [doi: 10.1111/1468-0009.12509] [Medline: 34075622]
30. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in outpatient care delivery and telemedicine during the COVID-19 pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
31. Notification of enforcement discretion for telehealth remote communications during the COVID-19 nationwide public health emergency. U.S. Department of Health & Humans Services. 2021 Jan 20. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html [accessed 2021-12-16]
32. Bonacini L, Gallo G, Scicchitano S. Working from home and income inequality: risks of a 'new normal' with COVID-19. J Popul Econ 2021 Sep 12;34(1):303-360 [FREE Full text] [doi: 10.1007/s00148-020-00800-7] [Medline: 32952308]
33. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
34. Outpatient telehealth use soared early in the COVID-19 pandemic but has since receded. Peterson-KFF Health System Tracker. 2022 Feb 10. URL: https://www.healthsystemtracker.org/brief/outpatient-telehealth-use-soared-early-in-the-covid-19-pandemic-but-has-since-receded/ [accessed 2022-07-24]
35. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? evidence and implications from a national sample of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
36. Turner K, Clary A, Hong Y, Alishahi Tabriz A, Shea CM. Patient portal barriers and group differences: cross-sectional national survey study. J Med Internet Res 2020 Sep 17;22(9):e18870 [FREE Full text] [doi: 10.2196/18870] [Medline: 32940620]
37. Stern RJ, Rafferty HF, Robert AC, Taniguchi C, Gregory B, Khoong EC, et al. Concentrating vaccines in neighborhoods with high COVID-19 burden. NEJM Catal Innov Care Deliv 2021 Apr 06 [FREE Full text]
38. Heath S. Was COVID-19 healthcare's use case for the patient portal? PatientEngagementHIT. 2020 Sep 18. URL: https://patientengagementhit.com/features/was-covid-19-healthcares-use-case-for-the-patient-portal [accessed 2022-07-24]
39. Craig S, Madu CR, Dalembert G. Never let a pandemic go to waste: applying an equity-focused quality improvement framework to close gaps in patient portal activation. Pediatrics 2021 Mar 01;147(3_MeetingAbstract):674-675 [FREE Full text]
40. 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 Jan 16;40(1):81-102. [doi: 10.1007/s11113-020-09634-3]
41. Rothbaum J. How does the pandemic affect survey response: using administrative data to evaluate nonresponse in the current population survey annual social and economic supplement. United States Census Bureau. 2021 Oct 08. URL: https://www.census.gov/newsroom/blogs/research-matters/2020/09/pandemic-affect-survey-response.html [accessed 2022-07-24]

## Abbreviations

**aOR:** adjusted odds ratio
**HINTS:** Health Information National Trends Survey

*Edited by C Basch; submitted 19.12.21; peer-reviewed by C Campos-Castillo, K Lawrence; comments to author 21.01.22; revised version received 02.05.22; accepted 26.08.22; published 19.09.22.*

*Please cite as:*
*Zeng B, Rivadeneira NA, Wen A, Sarkar U, Khoong EC*
*The Impact of the COVID-19 Pandemic on Internet Use and the Use of Digital Health Tools: Secondary Analysis of the 2020 Health Information National Trends Survey*
*J Med Internet Res 2022;24(9):e35828*
*URL: https://www.jmir.org/2022/9/e35828*
*doi:10.2196/35828*
*PMID:36041005*

Original Paper

# Engagement With HIV and COVID-19 Prevention: Nationwide Cross-sectional Analysis of Users on a Geosocial Networking App

Thomas W Gaither[1], MD, MAS; John T Sigalos[1], MD; Raphael J Landovitz[2,3], MD, MSc; Jesse N Mills[1], MD; Mark S Litwin[1,4,5], MD, MPH; Sriram V Eleswarapu[1], MD, PhD

[1]Department of Urology, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, United States

[2]Division of Infectious Diseases, Department of Medicine, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, United States

[3]Center for Clinical AIDS Research & Education, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, United States

[4]Department of Health Policy & Management, UCLA Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States

[5]School of Nursing, University of California, Los Angeles, Los Angeles, CA, United States

**Corresponding Author:**
Thomas W Gaither, MD, MAS
Department of Urology
David Geffen School of Medicine at UCLA
University of California, Los Angeles
10833 Le Conte Avenue, CHS 27-139
Los Angeles, CA, 90095
United States
Phone: 1 310 794 7700
Email: tgaither@mednet.ucla.edu

## Abstract

**Background:** Geosocial networking (GSN) apps play a pivotal role in catalyzing sexual partnering, especially among men who have sex with men.

**Objective:** To quantify the prevalence and disparities in disclosure of pre-exposure prophylaxis (PrEP) use and COVID-19 vaccination among GSN app users, mostly men who have sex with men, in the United States.

**Methods:** Web-based Grindr profiles from the top 50 metropolitan areas as well as the 50 most rural counties in the United States by population were randomly sampled. Grindr provides an option to disclose current PrEP use (HIV positive, HIV negative, or HIV negative with PrEP use). The free text in all profiles was analyzed, and any mention of COVID-19 vaccination was recorded. Multivariable logistic regression to assess independent associations with PrEP disclosure and COVID-19 vaccination was performed. Imputation analyses were used to test the robustness of the results.

**Results:** We evaluated 1889 urban and 384 rural profiles. Mean age among urban profiles was 32.9 (SD 9.6) years; mean age among rural profiles was 33.5 (SD 12.1) years ($P$=.41). Among the urban profiles, 16% reported being vaccinated against COVID-19 and 23% reported PrEP use compared to 10% and 8% in rural profiles, respectively ($P$=.002 and $P$<.001, respectively). Reporting COVID-19 vaccination (adjusted odds ratio [aOR] 1.7, 95% CI 1.2-2.4), living in an urban center (aOR 3.2, 95% CI 1.8-5.7), and showing a face picture as part of the Grindr profile (aOR 4.0, 95% CI 2.3-7.0) were positively associated with PrEP disclosure. Self-identified Black and Latino users were less likely to report PrEP use (aOR 0.6, 95% CI 0.4-0.9 and aOR 0.5, 95% CI 0.4-0.9, respectively). Reporting PrEP use (aOR 1.7, 95% CI 1.2-2.4), living in an urban center (aOR 2.5, 95% CI 1.4-4.5), having a “discreet” status (aOR 1.6, 95% CI 1.0-2.5), and showing a face picture (aOR 2.7, 95% CI 1.5-4.8) were positively associated with reporting COVID-19 vaccination on their profile. Users in the southern United States were less likely to report COVID-19 vaccination status than those in the northeast United States (aOR 0.6, 95% CI 0.3-0.9).

**Conclusions:** Variations in PrEP disclosure are associated with race, whereas COVID-19 vaccination disclosure is associated with geographic area. However, rural GSN users were less likely to report both PrEP use and COVID-19 vaccination. The data demonstrate a need to expand health preventative services in the rural United States for sexual minorities. GSN platforms may be ideal for deployment of preventative interventions to improve access for this difficult-to-reach population.

***(J Med Internet Res 2022;24(9):e38244)*** doi:10.2196/38244

**KEYWORDS**

geosocial networking apps; pre-exposure prophylaxis; vaccination; rural; men who have sex with men; surveillance; digital surveillance; COVID-19; digital application; geosocial network; public health; surveillance platform; health platform; mobile health

## *Introduction*

Pre-exposure prophylaxis (PrEP) is a highly effective and safe strategy to prevent the acquisition of HIV [1-3]. The efficacy of oral PrEP is dependent on adherence; a long-acting injectable PrEP agent was recently shown to be superior to daily oral PrEP and received US Food and Drug Administration approval on December 21, 2021 [4]. Despite these advances, men who have sex with men continue to represent 65% of new HIV diagnoses in 2019 in the United States [5]. Among this group, Black or African American and Hispanic or Latino men who have sex with men are disproportionately affected, especially within the southern United States [5]. These disparities in HIV transmission reflect many complex, psychosocial phenomena, including insular networks, social and HIV-related stigma, differential access to medical services, poverty, housing instability, racism, victimization, trauma, and discrepant PrEP awareness, access, and advocacy [6,7]. Although public disclosure of PrEP use within the men who have sex with men community may serve as a mechanism of educating others and serve as confirmation of an HIV-negative status, concerns about stigma or being labeled as sexually promiscuous limit widespread disclosure [8].

Lifetime usage of geosocial networking (GSN) apps among men who have sex with men is increasingly common, with Grindr being the frequently used by this population in the United States [9]. In a survey of adolescent men who have sex with men, 53% of them reported using apps specific to this population to meet partners for sexual intercourse [10]. Grindr allows users to locate other users within close proximity and is recognized as a “hookup” app, indicating that users seek casual, often anonymous sexual partners. More than twice as many Grindr-using men who have sex with men report 5 or more partners within the past 12 months than those who do not use the app (60% vs 27%, respectively) [11]. Many GSN apps, including Grindr, allow disclosure of HIV and PrEP status on user profiles, which may contribute to sero-sorting behaviors [12]. In fact, many users judge trustworthiness and perceived risk by observing others’ profile characteristics. Lack of pictures, minimal bios, grammatically incorrect bios, or lack of linked social media accounts are viewed as less trustworthy [13]. In a nationwide survey of urban Grindr profiles, 18.1% of users reported PrEP use, and 61% reported their HIV status [14]. PrEP disclosure was 40% less common among Black users than among White users and 30% less common among users in the southern United States than in the northeastern United States [14]. These disparities in PrEP disclosures on GSN apps mirror those reported by the Centers for Disease Control and Prevention (CDC) and, as such, may provide a real-world avenue for epidemiological surveillance.

Since the emergence of SARS-CoV-2 (the virus causing COVID-19) in 2019, HIV prevention and sexually transmitted infection (STI)–testing services have been disrupted [15]. Additionally, the pandemic has impacted sexual behavior by mostly reducing the number of sexual partners among men who have sex with men and has caused some to self-discontinue PrEP entirely or to selectively skip doses [16,17]. However, many men who have sex with men did not change sexual practices during the pandemic [18]. Reported rates of condomless anal sex with casual partners have declined; however, the association between drug use and risky sexual behavior increased [19]. Despite variations in overall sexual activity, the pandemic may be interrupting access to antiretroviral therapy, which has implications for HIV transmission [20,21]. Since implementation of social distancing and widespread closures of bars and clubs, usage of GSN apps specifically among men who have sex with men has increased, and 37% of men who have sex with men who participated in a web-based survey (N=2562) reported spending more time on apps searching for sexual partners than in prepandemic times [18]. Others have turned to GSN apps for social connectivity during lonely quarantine periods [22]. Overall, the COVID-19 pandemic has shifted sexual behavior, and GSN apps appear to play a pivotal role in catalyzing sexual partnering.

Any close contact, including but not limited to sexual intercourse, increases the risk of COVID-19 infection; vaccination against SARS-CoV-2 mitigates this risk [23,24]. How vaccination impacts the selection of sexual partners among the community of men who have sex with men remains understudied. In response to the intersection of the COVID-19 and HIV pandemics, we aimed to assess health promotional strategies, including PrEP use and COVID-19 vaccination, in web-based in a national sample of Grindr users. We sought to determine geographical and demographic trends in PrEP use and COVID-19 vaccination status disclosure to suggest geographic areas for targeting health promotion interventions. We hypothesized that men who have sex with men in rural areas would report less use of PrEP and lower disclosure of vaccination against COVID-19.

## *Methods*

### Sample Population

Grindr profiles from the 50 most populated metropolitan areas as well as the 50 least populated rural counties in the United States were randomly sampled [25,26]. Data collection occurred from September through December 2021 between 3 PM and 8 PM local time to ensure standardization of data collection and to eliminate time of day as a confounding factor. Each city or rural county was sampled in a random order determined by a random number generator. The number of profiles sampled from each city was proportional to the percentage of reported individuals identifying as lesbian, gay, bisexual, or transgender within that city [25]. Rural county users represented 17% of the sample, which reflects the US population [27]. Profiles were collected in the order in which they appeared within the city center or the geographic center of the rural county. A profile was created to collect the data. The profile was blank, did not contain any information, and no users were messaged. Based on previous research, we expect 18% of urban users to report PrEP use [14]. If we estimate a 6% difference as clinically significant between urban and rural users, we estimated a need to sample 2100 profiles with an 80% power and a 2-tailed α of .05.

### Exposures

All information provided within the profile was systematically abstracted into a REDCap database independently by 2 authors (TWG and JTS). REDCap is a browser-based electronic data capture software and workflow to organize and store clinical and translational research data. All geographic locations were categorized into US census region (Northeast, Midwest, South, and West) and as urban versus rural regions [27]. Available demographic characteristics included age, race and ethnicity, BMI (calculated from height and weight), gender, and relationship status. Variables were collected as presented on the app and are all self-reported. Race and ethnicity were categorized as Asian, Black, Latino, Middle Eastern, Mixed, Native American, White, South Asian, other; gender as man, cisgender man, transgender man, woman, transgender woman, nonbinary, nonconforming, queer, and crossdresser; and relationship status as single, dating, exclusive, committed, partnered, engaged, married, and open relationship. Owing to small numbers, gender was dichotomized as *cisgender male* or *gender diverse*, and relationship status was similarly dichotomized as *single* or *relationship*. Irrespective of whether the user profile presented a face picture, a body picture, other photos (neither face nor body), or no picture (blank) were abstracted. Picture status was considered a surrogate for *willingness to share*, as those who share a face picture have been found to be associated with a likelihood of reporting more data about themselves, including PrEP use [14,28].

### Outcomes

Disclosure of PrEP and vaccination status were the primary outcomes of interest. Grindr provides an option to disclose current PrEP use (HIV positive, HIV negative, or HIV negative and on PrEP). Profiles not reporting PrEP use were assumed not to be taking PrEP. The most common reason why users do not report PrEP status on GSN profiles is that they are not taking PrEP [12]. Grindr does not provide an option to report COVID-19 vaccination status. However, all free text of profiles were read and any mention of being vaccinated against COVID-19 was recorded. No profiles explicitly stated that they were not vaccinated. As such, both PrEP and vaccination outcomes in our study may represent *engagement* with HIV or COVID-19 prevention rather than actual PrEP use or vaccination status. Date of the last reported STI or HIV testing was recorded when available (n=934). Thus, our secondary outcome was the calculated time between data collection and the last STI or HIV test (in months).

### Other Variables

Other available self-reported variables were abstracted, including body type (toned, average, large, muscular, slim, or stocky), sexual position (bottom, versatile bottom, versatile, versatile top, or top), app use intent (chats, dates, friends, networking, relationship, or “right now”), and gay tribe, which is representation of certain subcultures (bear, clean-cut, daddy, discreet, jock, leather, otter, poz, rugged, trans, twink, and sober). .

### Statistical Analysis

Demographic characteristics between urban and rural profiles were compared to understand statistical correlates and to describe the population. Multivariable logistic regression was used to assess the independent factors associated with PrEP use and COVID-19 vaccination. Covariates were statistically selected ($P<.01$) on the basis of bivariate analysis. Age was included in all models owing to the availability of this variable. Because age, BMI, and race and ethnicity were not available in all profiles, the robustness of our base model was assessed with 2 imputed models—this was done to prevent the exclusion of observed data and estimate the range of selection bias [29]. The base model included only nonmissing data.

### Imputed Model A

Stepwise linear regression was used with all known variables within the data set to predict age and BMI. Actual age and BMI values were used preferentially over imputed values. Race and ethnicity were imputed using the previous known value. This approach assumes data were missing at random. A distribution of the imputed values in model A is provided in Multimedia Appendix 1.

### Imputed Model B

Data are unlikely to be missing at random. For example, older users, those with overweight or obesity, and racial minorities are more likely to experience discrimination in social environments on the internet and thus may omit this information intentionally [30]. To understand the uncertainty in the data, we used the same aforementioned imputation model but added 10 years of age to all imputed age values and +5 units to all imputed BMI values. Profiles with missing race and ethnicity data were imputed to the most common racial minority within the US census region of where the profile was found [27]. A distribution of the imputed values in model B is provided in Multimedia Appendix 1.

For our secondary outcome, the number of months since the users’ last STI or HIV test was compared among the demographic and sexual characteristics collected. Because the number of months since last testing was not normally distributed, Mann-Whitney *U* tests were used for bivariate comparisons and Kruskal-Wallis tests were performed for nominal variables. All statistical tests were 2-sided, and $P<.05$ was considered statistically significant in the final multivariable models. Analyses were completed in Stata (version 17; StataCorp).

### Ethics Approval

The institutional review board of the University of California, Los Angeles exempted the study from review as no participant protected health information was collected and no contact with users via their profiles was made.

## Results

We randomly sampled 1889 urban and 384 rural profiles. Table 1 shows the comparison of demographics by urban versus rural status. The mean age among urban profiles was 32.9 (SD 9.6) years and that among rural profiles was 33.5 (SD 12.1) years (*P*=.41). Consistent with census data, 84% of rural counties were concentrated in the western United States, whereas urban centers were distributed across US regions, (*P*<.001). Among the urban profiles, 296 (16%) reported being vaccinated against COVID-19 and 426 (23%) reported PrEP use compared to 37 (10%) and 30 (8%) in rural profiles, respectively (*P*=.002 and *P*<.001, respectively). Black users comprised 15% of urban users compared to 4% in rural areas, and Native American users comprised 0.5% in urban areas compared to 7% in rural areas. Urban users had lower BMIs than rural users (25.2 vs 27.2; *P*<.001). There were no statistically significant differences between urban and rural profiles regarding gender, relationship status, HIV disclosure status, or sexual position preference. On average, urban profiles reported STI or HIV testing within approximately 4 months versus 8 months in rural profiles (*P*<.001). Blank and “other” photos were more common among rural profiles than among urban profiles (45% vs 26%; *P*<.001).

**Table 1.** Demographic and sexual characteristics of urban versus rural Grindr profiles in the United States, 2021.

| | Urban profiles (n=1889) | Rural profiles (n=384) | *P* value | Missing data in the data set, % |
|---|---|---|---|---|
| **Demographic characteristics** | | | | |
| Age (years), mean (SD)[a] | 32.9 (9.6) | 33.5 (12.1) | .41 | 18.3 |
| **Location in the United States, n (%)** | | | <.001 | 0 |
| Northeast | 403 (21) | 0 (0) | | |
| Midwest | 355 (19) | 22 (6) | | |
| South | 645 (34) | 41 (11) | | |
| West | 486 (26) | 321 (84) | | |
| Included vaccination status on profile, n (%) | 296 (16) | 37 (10) | .002 | 0 |
| Taking pre-exposure prophylaxis, n (%) | 426 (23) | 30 (8) | <.001 | 0 |
| **Ethnicity, n (%)[a]** | | | <.001 | 28.9 |
| Asian | 71 (5) | 1 (2) | | |
| Black | 203 (15) | 11 (4) | | |
| Latino | 231 (17) | 38 (14) | | |
| Middle Eastern | 15 (1) | 0 | | |
| Mixed | 117 (9) | 27 (10) | | |
| Native American | 4 (0.5) | 18 (7) | | |
| White | 682 (51) | 164 (61) | | |
| South Asian | 6 (0.5) | 1 (0.5) | | |
| Other | 20 (1) | 3 (1) | | |
| BMI, mean (SD)[a] | 25.2 (3.9) | 27.2 (4.2) | <.001 | 28.7 |
| **Gender, n (%)[a]** | | | .73 | 39.9 |
| Cisgender male | 1051 (94) | 234 (94) | | |
| Gender diverse | 65 (6) | 16 (6) | | |
| **Relationship status, n (%)[a]** | | | .54 | 41.9 |
| Single | 883 (81) | 185 (79) | | |
| Relationship | 205 (19) | 48 (21) | | |
| Time until last test for sexually transmitted infections (months), median (IQR)[a] | 4 (2-9) | 8 (3-14) | <.001 | 58.9 |
| Indicated HIV status on profile, n (%) | 1148 (61) | 213 (55) | .05 | 0 |
| **Picture status, n (%)** | | | <.001 | 0 |
| Face picture | 1000 (53) | 150 (39) | | |
| Body picture | 399 (21) | 60 (16) | | |
| Random photo | 88 (5) | 59 (15) | | |
| Blank | 402 (21) | 115 (30) | | |
| **Sexual characteristics** | | | | |
| **Sexual position, n (%)[a]** | | | .24 | 34.1 |
| Bottom | 211 (18) | 58 (20) | .24 | |
| Vers bottom | 146 (12) | 38 (13) | | |
| Vers | 384 (32) | 104 (36) | | |
| Vers top | 178 (15) | 37 (13) | | |
| Top | 287 (24) | 54 (19) | | |

| | Urban profiles (n=1889) | Rural profiles (n=384) | *P* value | Missing data in the data set, % |
|---|---|---|---|---|
| **Grindr tribe, n (%)** | | | | 0 |
| Bear | 88 (5) | 29 (8) | .02 | |
| Clean-cut | 172 (9) | 42 (11) | .26 | |
| Daddy | 143 (8) | 34 (9) | .39 | |
| Discreet | 236 (12) | 71 (18) | .002 | |
| Geek | 120 (6) | 18 (5) | .21 | |
| Jock | 177 (9) | 27 (7) | .14 | |
| Leather | 32 (2) | 4 (1) | .35 | |
| Otter | 87 (5) | 16 (4) | .71 | |
| Poz | 11 (1) | 1 (0) | .43 | |
| Rugged | 55 (3) | 22 (6) | .01 | |
| Trans | 60 (3) | 14 (4) | .64 | |
| Twink | 103 (5) | 14 (4) | .14 | |
| Sober | 15 (1) | 4 (1) | .63 | |
| **Response to *Looking for*, n (%)** | | | | 0 |
| Chats | 749 (40) | 167 (43) | .16 | |
| Dates | 658 (35) | 118 (31) | .12 | |
| Friends | 875 (46) | 184 (48) | .57 | |
| Networking | 360 (19) | 73 (19) | .98 | |
| Relationship | 458 (24) | 86 (22) | .44 | |
| Right now | 952 (50) | 223 (58) | .01 | |
| **Response to *Meet at*, n (%)** | | | | 0 |
| My place (host) | 451 (24) | 96 (25) | .64 | |
| Your place (travel) | 582 (31) | 145 (38) | .01 | |
| Bar | 299 (16) | 60 (16) | .92 | |
| Coffee shop | 295 (16) | 65 (17) | .52 | |
| Restaurant | 246 (13) | 52 (14) | .78 | |

[a] Missing data excluded.

There were no significant differences in most Grindr tribes among urban versus rural users. There were more “discreet” users in rural areas than in urban areas (18% vs 12%; *P*=.002). There were more rural users who were using the app for the intent of immediate sexual partnering (ie, “right now”) compared to urban users (58% vs 50%; *P*=.01). Additionally, more rural users indicated a preference to meet at someone else’s location than did urban users (38% vs 31%; *P*=.01).

Table 2 shows the results of the multivariable model and imputation analyses. Reporting COVID-19 vaccination status (adjusted odds ratio [aOR] 1.7, 95% CI 1.2-2.4), living in an urban center (aOR 3.2, 95% CI 1.8-5.7), and showing a face picture (aOR 4.0, 95% CI 2.3-7.0) were positively associated with PrEP use. Self-identified Black and Latino users were less likely to report PrEP use (aOR 0.6, 95% CI 0.4-0.9 and aOR 0.5, 95% CI 0.4-0.9, respectively). All of these associations remained in the sensitivity analyses. Point estimates in the base model and model A of the association between BMI and PrEP use were not statistically significant; however, in model B, users with increasing BMI were negatively associated with PrEP use (aOR 0.85, 95% CI 0.74-0.97).

**Table 2.** Multivariable analysis of raw data and imputation models of engagement with the use of pre-exposure prophylaxis (PrEP) and COVID-19 vaccination indicated on Grindr profiles, 2021.

| | PrEP use | | | COVID-19 vaccination status | | |
|---|---|---|---|---|---|---|
| | PrEP use (n=1107), adjusted odds ratio (aOR) (95% CI) | Imputed model A[a] (n=2273), aOR (95% CI) | Imputed model B[b] (n=2273), aOR (95% CI) | COVID-19 vaccination (n=1084), aOR (95% CI) | Imputed model A[a] (n=2255), aOR (95% CI) | Imputed model B[b] (n=2258), aOR (95% CI) |
| Age (every 10 years) | 0.9 (0.8-1.1) | 1.0 (0.9-1.1) | 1.0 (0.9-1.1) | 1.1 (0.9-1.3) | 1.2 (1.0-1.3)[c] | 1.2 (1.1-1.4)[c] |
| PrEP use | —[d] | — | — | 1.7 (1.2-2.4)[c] | 1.8 (1.4-2.4)[c] | 1.9 (1.4-2.5)[c] |
| Vaccinated | 1.7 (1.2-2.4)[c] | 1.9 (1.4-2.4)[c] | 1.9 (1.4-2.5)[c] | — | — | — |
| Living in urban centers | 3.2 (1.8-5.7)[c] | 2.9 (1.9-4.4)[c] | 2.8 (1.8-4.3)[c] | 2.5 (1.4-4.5)[c] | 2.2 (1.5-3.4)[c] | 2.2 (1.4-3.3)[c] |
| **Region** | | | | | | |
| Northeast | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) |
| Midwest | 0.9 (0.5-1.4) | 0.8 (0.6-1.2) | 0.8 (0.6-1.2) | 1.1 (0.6-1.8) | 1.3 (0.9-2.0) | 1.2 (0.8-1.8) |
| South | 0.8 (0.5-1.2) | 1.0 (0.7-1.3) | 1.0 (0.8-1.4) | 0.6 (0.3-0.9)[c] | 0.7 (0.4-0.9)[c] | 0.6 (0.4-0.9)[c] |
| West | 0.8 (0.5-1.3) | 0.9 (0.7-1.2) | 0.9 (0.7-1.3) | 1.3 (0.8-2.1) | 1.5 (1.0-2.2)[c] | 1.5 (1.0-2.2)[c] |
| **Race and ethnicity** | | | | | | |
| Asian | 1.0 (0.5-1.8) | 1.2 (0.7-1.8) | 0.9 (0.6-1.6) | 0.7 (0.3-1.7) | 1.1 (0.6-1.9) | 1.2 (0.6-2.2) |
| Black | 0.6 (0.4-0.9)[c] | 0.6 (0.4-0.9)[c] | 0.6 (0.4-0.8)[c] | 1.3 (0.8-2.1) | 1.0 (0.7-1.5) | 1.1 (0.8-1.6) |
| Latino | 0.6 (0.4-0.9)[c] | 0.7 (0.5-0.9)[c] | 0.7 (0.4-0.9)[c] | 0.8 (0.5-1.4) | 0.9 (0.6-1.3) | 0.8 (0.6-1.1) |
| Middle eastern | 0.4 (0.1-2.0) | 0.9 (0.3-2.8) | 0.6 (0.2-2.3) | —[e] | —[e] | —[e] |
| Mixed | 0.8 (0.5-1.3) | 0.9 (0.6-1.3) | 0.8 (0.5-1.2) | 1.2 (0.7-2.0) | 1.0 (0.6-1.5) | 1.1 (0.7-1.8) |
| Native American | 1.4 (0.4-5.9) | 0.9 (0.3-2.5) | 0.9 (0.2-3.3) | 1.1 (0.2-5.3) | 1.3 (0.5-3.4) | 1.5 (0.5-4.7) |
| South Asian | 0.6 (0.1-6.6) | 0.4 (0.1-2.0) | 0.4 (0-3.1) | 1.6 (0.2-16) | 2.0 (0.5-7) | 2.7 (0.5-15) |
| Other | 1.7 (0.4-6.3) | 1.5 (0.7-3.5) | 2.0 (0.8-5.0) | — | 0.3 (0-1.4) | 0.2 (0-1.4) |
| White | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) |
| BMI (every 5 points) | 1.0 (0.8-1.2) | 1.0 (0.8-1.1) | 0.9 (0.7-0.9)[c] | 1.2 (0.9-1.4) | 1.3 (1.1-1.5)[c] | 1.2 (1.0-1.3)[c] |
| Discreet | 1.0 (0.7-1.6) | 1.2 (0.9-1.7) | 1.2 (0.8-1.6) | 1.6 (1.0-2.5)[c] | 1.6 (1.1-2.3)[c] | 1.6 (1.1-2.3)[c] |
| **Picture status** | | | | | | |
| Face | 4.0 (2.3-7.0)[c] | 3.7 (2.6-5.2)[c] | 3.3 (2.3-4.7)[c] | 2.7 (1.5-4.8)[c] | 3.2 (2.1-4.7)[c] | 3.2 (2.2-4.8)[c] |
| Body only | 2.3 (1.2-4.4)[c] | 2.3 (1.5-3.4)[c] | 2.1 (1.4-3.1)[c] | 1.6 (0.8-3.2) | 2.2 (1.4-3.4)[c] | 2.2 (1.4-3.4)[c] |
| Random photo | 2.0 (0.8-5.5) | 0.9 (0.4-1.8) | 0.8 (0.4-1.7) | 1.8 (0.7-5.0) | 2.0 (1.1-3.8)[c] | 1.9 (1.0-3.6)[c] |
| Blank | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) | 1.0 (referent) |

[a] Imputed model was generated through stepwise regression of known values (Multimedia Appendix 1); race and ethnicity were imputed on the basis of previous value, assuming data are missing at random.

[b] Imputed model was generated with nonrandom missing values; imputed age+10, imputed BMI +5, and race and ethnicity imputed to most common racial minority within the US census region.

[c] Boldfaced values indicate significant values at *P*<.05.

[d] Not determined.

[e] In the imputed model, none of the 15 Middle Eastern users were vaccinated and thus excluded from the model.

Reporting PrEP use (aOR 1.7, 95% CI 1.2-2.4), living in an urban center (aOR 2.5, 95% CI 1.4-4.5), having a "discreet" status (aOR 1.6, 95% CI 1.0-2.5), and showing a face picture (aOR 2.7, 95% CI 1.5-4.8) were positively associated with report of COVID-19 vaccination status on their profiles. Users in the southern United States were less likely to report being vaccinated (aOR 0.6, 95% CI 0.3-0.9) than those in the northeast United States. These associations remained in the imputation models. In both imputation models, users in the western United States were more likely to report being vaccinated (aOR 1.5, 95% CI 1.0-2.2) than those in the northeast United States, and users with increasing BMI were more likely to report vaccination (aOR 1.6, 95% CI 1.1-2.3). Model accuracy information is summarized in Table 3. Overall, the model is highly specific and poorly sensitive. Accuracy ranged from 73.9% to 85.3%.

**Table 3.** Accuracy of the models.

| | Use of pre-exposure prophylaxis (PrEP) | | | COVID-19 vaccination status | | |
|---|---|---|---|---|---|---|
| | PrEP use (n=1107) | Imputed model A (n=2273) | Imputed model B (n=2273) | COVID-19 vaccination (n=1084) | Imputed model A (n=2255) | Imputed model B (n=2258) |
| Sensitivity, % | 0.7 | 0 | 0.4 | 1.1 | 0.9 | 0.6 |
| Specificity, % | 98.7 | 100 | 99.5 | 99.9 | 99.9 | 99.9 |
| Correctly classified, % | 73.9 | 79.9 | 79.6 | 82.5 | 85.3 | 85.3 |

Table 4 shows the time (in months) since the last STI or HIV test. The median time since the last STI or HIV test for the whole sample was 4 (IQR 2-10) months. Notable differences in the timing of the last STI or HIV test include the following: living in an urban versus rural center (4 months vs 8 months; *P*<.001); reporting versus not reporting PrEP use (3 months vs 6 months; *P*<.001), and BMI category (4 months: users with normal weight, 5 months: users with overweight, and 6 months: users with obesity; *P*=.003).

**Table 4.** Time since last test for sexually transmitted infections or HIV in our nationwide sample of Grindr users, 2021.

| | Time (months), median (IQR) | *P* value[a] |
|---|---|---|
| **Age group (years)[b]** | | .03 |
| ≤30 | 4 (2-9) | |
| >30 | 4 (2-11) | |
| **Urbanicity** | | <.001 |
| Urban | 4 (2-9) | |
| Rural | 8 (3-14) | |
| **Location in the United States** | | .53 |
| Northeast | 4 (2-8) | |
| Midwest | 4 (2-10) | |
| South | 4 (2-9) | |
| West | 5 (2-11) | |
| **COVID-19 vaccination status** | | .02 |
| Yes | 3 (2-7) | |
| No | 4 (2-11) | |
| **Status of use of pre-exposure prophylaxis** | | <.001 |
| Yes | 3 (1-5) | |
| No | 6 (3-13) | |
| **Ethnicity** | | .35 |
| Asian | 5 (2-12) | |
| Black | 3 (2-7) | |
| Latino | 6 (2-12) | |
| Middle Eastern | 6 (4-10) | |
| Mixed | 4 (2-10) | |
| Native American | 4 (2-10) | |
| White | 4 (2-10) | |
| South Asian | 5 (5-9) | |
| Other | 13 (4-15) | |
| **BMI** | | .003 |
| Normal weight | 4 (2-9) | |
| Overweight | 5 (2-10) | |
| Obese | 6 (3-16) | |
| **Gender** | | .01 |
| Cisgender male | 4 (2-10) | |
| Gender diverse | 2 (1-7) | |
| **Relationship status** | | .35 |
| Single | 4 (2-10) | |
| In a relationship | 4 (2-11) | |
| **HIV status on profile** | | .18 |
| Yes | 4 (2-10) | |
| No | 6 (2-12) | |
| **Picture status** | | .10 |

| | Time (months), median (IQR) | *P* value[a] |
|---|---|---|
| Face picture | 4 (2-10) | |
| Body picture | 4 (2-9) | |
| Random photo | 7 (3-14) | |
| Blank | 4 (2-9) | |
| **Sexual position** | | .10 |
| Bottom | 4 (2-9) | |
| Vers bottom | 3 (2-8) | |
| Vers | 4 (2-10) | |
| Vers top | 4 (2-9) | |
| Top | 5 (2-12) | |

[a] All *P* values are nonparametric tests owing to the nonnormal distribution of time since the last test for sexually transmitted infection.

[b] 31 years is the median age of the sample.

## Discussion

### Principal Findings

This study provides comparative estimates of engagement in preventative health within a large nationwide sample of individuals in the United States, mostly men who have sex with men, on one GSN app during the COVID-19 pandemic. The study highlights several important differences in infection prevention behaviors between urban and rural app users. After adjustment, urban users are approximately 3 times as likely to disclose PrEP usage (aOR 3.2) and approximately 2.5 times as likely to report being vaccinated against COVID-19 (aOR 2.5). PrEP use and vaccination were often co-reported. Additionally, rural users' last self-reported STI or HIV test was approximately 4 months delayed compared to that among urban users. Race and ethnicity appear to be highly associated with PrEP disclosure, whereas geography but not race or ethnicity appear to be associated with reporting of COVID-19 vaccination status.

Rural men who have sex with men experience substantial stigma and lack robust access to HIV preventative care [31-34]. Our results corroborate this experience in a real-world setting of users on GSN app specific to men who have sex with men, and demonstrate the need to expand preventive health services to these populations—in fact, the digital fluency of these users in leveraging a GSN-based app for sexual partnership suggests that such platforms may be ideal for deployment of interventions to support preventive care among rural populations of men who have sex with men. Internet-based HIV preventative messaging to rural men who have sex with men has been shown to be acceptable and efficacious [35]. Large-scale implementation of such programs is lacking. GSN apps may provide not only avenues for understanding spatial distribution of rural men who have sex with men but also provide a private arena to deliver care, which has been reported as an acceptable platform to deliver health information and services [36-38].

Rural users were more likely to not show their faces on their profiles and to report being "discreet." Men who identify as discreet primarily seek sexual partners on the internet as it is seen as socially safer and can provide anonymity [39]. Potential reasons for discretion include stigma, internalized homophobia, or being in a relationship [39]. HIV-related stigma may increase the likelihood of risky sexual behavior and mental health or substance use disorders [40]. Only 8% of rural users in our study reported PrEP use, and the median time since the last STI or HIV test among rural users was 8 months. This may reflect barriers to STI HIV testing during the pandemic and more durable limitations to HIV preventative services in these areas. The rise of telemedicine access during the pandemic may serve to enfranchise the historically underserved rural population of men who have sex with men, who may benefit from improved web-based STI and HIV prevention services [33]. In response to the pandemic, many clinics for sexual minorities have developed telemedicine infrastructure for continued STI or HIV prevention services [41,42]. Removing barriers such as access and privacy concerns may improve screening and increase PrEP uptake among rural men who have sex with men [43].

Despite controlling for age, region, living in urban versus rural centers, BMI, picture status, and discreetness, Black and Latino users were less likely to report PrEP use. In the National HIV Behavioral Surveillance data analyzed by the CDC, only 30% of Hispanic and 26% of Black men who have sex with men reported taking PrEP within the past year compared to 42% of White men who have sex with men [44]. Similar disparities were observed in the southern United States compared to other regions independent of race and ethnicity [44]. PrEP is exclusively available via prescription except within states that allow pharmacists to provide this service [45]. Barriers to PrEP for Black or Latino men who have sex with men include racism, provider bias, insurance coverage, and marginalization [46,47]. The southern United States, as a region, has unique challenges to PrEP uptake, including a geographically expansive rural population, low rates of commercial insurance coverage, lack of Medicaid expansion, social stigma, low health literacy, and low HIV risk perception [48].

Despite the COVID-19 pandemic, the overall prevalence of PrEP use among urban users increased since 2018 (18% to 23%) [14]. Worldwide survey evidence suggests that many men who have sex with men have decreased or completely stopped taking PrEP during the pandemic [16,49-51]. Discontinuing PrEP during sexually inactive periods and resuming PrEP during activities that increase the risk of HIV exposure has been termed “preventive-effective” adherence [52]. Thus, some men who have sex with men who use this strategy may still report PrEP use on their GSN profile. It is plausible that the pandemic brought more PrEP-using men who have sex with men to GSN apps to meet partners after the closure of social gathering spaces such as bars and clubs. An alternative explanation is that app users who discontinued PrEP during the pandemic did not change their PrEP status on their profile. Growing evidence suggests that PrEP status can be viewed as a surrogate for desiring condomless anal sex, and some users do not present accurate information on the internet [53,54].

Whether the veracity of health-related information on GSN apps should be verified merits further inquiry. In 2011, the adult film industry created the Performer Availability Screening Service (PASS) program, which is a database to show “work clearance” by documenting STI and HIV status. The PASS program now includes verification of COVID-19 vaccination status. Widespread implementation of a system such as the PASS program would require public health and industry collaboration as well as user buy-in.

The prevalence of COVID-19 vaccination reporting within the whole sample was low (333/2273, 14.65%). The likely main driver of the overall low prevalence is that vaccination status was abstracted from free text and therefore passively reported, in contrast to HIV or PrEP status. Actual vaccination rates have been shown to be higher among sexual minorities than among heterosexual adults (85.4% vaccinated among gay and lesbian adults vs 76.3% of heterosexual adults) [55]. Interestingly, our results show that vaccination reporting is more closely associated with geographic status rather than with race or ethnicity. Lower uptake of the vaccine in rural areas is consistent with CDC data [56]. Vaccine willingness has also been associated with political leaning, which is similarly geographically associated [57]. Although Black race has been associated with lower vaccine acceptance, this was not supported by our data [58]. However, our study may be underpowered to detect these differences. If high rates of vaccine acceptability or uptake can indeed be confirmed in this population, it may provide supportive data to advocate for vaccine-based STI prevention strategies, such as the ongoing MAGI trial leveraging the serogroup B meningococcal vaccine as a potential *Neisseria gonorrhoeae* vaccine (ClinicalTrials.gov NCT04350138) [59,60]. Our data provide further evidence that once structural barriers are decreased, public health interventions may be well received among users of GSN apps.

## Limitations

Our study has certain limitations. Users of GSN apps, especially during the pandemic, may not be generalizable to all men who have sex with men. We intentionally did not sample users in suburban environments to contrast the stark geographical effects between urban and rural users. As the pandemic has shifted sexual behavior, not all users are sexually active. Important confounding variables such as education and income, which strongly impact health care access and uptake, were not available in the data set, and indeed these may be central drivers of the observed findings. All variables are self-reported and may be prone to social desirability bias. Our main outcome of interest likely represents engagement with health prevention rather than actual PrEP use or COVID-19 vaccination. We compared users who disclosed to those who did not disclose PrEP use or COVID-19 vaccination status. Young Black men who have sex with men in the southern United States report high usage of GSN apps to meet sexual partners, but Grindr was the least commonly used app [61]. The rates of PrEP disclosure among Black or Latino men who have sex with men who use other apps remain to be studied. Imputation analyses are dependent on the reasons for missing data and can significantly impact measurement error, limiting power. Overall, the modeling accuracy was adequate, but the sensitivity was extremely poor. The statistical aim of the modeling was to assess associations with our outcomes of interest and not to generate a prediction model.

## Conclusions

We demonstrate a need to expand health preventative services for sexual minorities in the rural United States. Disclosure of PrEP use and COVID-19 vaccination status on GSN app profiles mirror trends assessed by CDC survey data. Variations in PrEP disclosure are largely impacted by race, whereas COVID-19 vaccination disclosure is influenced by geographic area; however, rural GSN users were less likely to report both PrEP use and COVID-19 vaccination status. GSN platforms may be ideal for deployment of preventative interventions to improve access for this difficult-to-reach population.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental table.
[DOCX File , 19 KB - jmir_v24i9e38244_app1.docx ]

## References

1. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599. [doi: 10.1056/nejmoa1011205]
2. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med 2012 Aug 02;367(5):399-410. [doi: 10.1056/nejmoa1108524]

3. Mayer KH, Molina J, Thompson MA, Anderson PL, Mounzer KC, De Wet JJ, et al. Emtricitabine and tenofovir alafenamide vs emtricitabine and tenofovir disoproxil fumarate for HIV pre-exposure prophylaxis (DISCOVER): primary results from a randomised, double-blind, multicentre, active-controlled, phase 3, non-inferiority trial. The Lancet 2020 Jul;396(10246):239-254. [doi: 10.1016/s0140-6736(20)31065-5]
4. Landovitz RJ, Donnell D, Clement ME, Hanscom B, Cottle L, Coelho L, et al. Cabotegravir for HIV prevention in cisgender men and transgender women. N Engl J Med 2021 Aug 12;385(7):595-608. [doi: 10.1056/nejmoa2101016]
5. HIV surveillance reports. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2022-08-29]
6. Ezennia O, Geter A, Smith DK. The PrEP care continuum and Black men who have sex with men: a scoping review of published data on awareness, uptake, adherence, and retention in PrEP care. AIDS Behav 2019 Oct 28;23(10):2654-2673. [doi: 10.1007/s10461-019-02641-2] [Medline: 31463711]
7. Mustanski B, Morgan E, D Aquila R, Birkett M, Janulis P, Newcomb M. Individual and network factors associated with racial disparities in HIV among young men who have sex with men: results from the RADAR cohort study. J Acquir Immune Defic Syndr 2019 Jan 01;80(1):24-30 [FREE Full text] [doi: 10.1097/QAI.0000000000001886] [Medline: 30365451]
8. Bartels S, Castillo I, Davis DA, Hightow-Weidman LB, Muessig KE, Galindo C, et al. PrEP disclosure experiences of gay and bisexual men in Guatemala. AIDS Behav 2021 Dec 08;25(12):4115-4124. [doi: 10.1007/s10461-021-03276-y] [Medline: 33966139]
9. Goedel WC, Duncan DT. Geosocial-networking app usage patterns of gay, bisexual, and other men who have sex with men: survey among users of Grindr, a mobile dating app. JMIR Public Health Surveill 2015 May 08;1(1):e4 [FREE Full text] [doi: 10.2196/publichealth.4353] [Medline: 27227127]
10. Macapagal K, Moskowitz DA, Li DH, Carrión A, Bettin E, Fisher CB, et al. Hookup App Use, Sexual Behavior, and Sexual Health Among Adolescent Men Who Have Sex With Men in the United States. J Adolesc Health 2018 Jun;62(6):708-715 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.01.001] [Medline: 29784114]
11. Landovitz RJ, Tseng C, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug 15;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]
12. Medina MM, Crowley C, Montgomery MC, Tributino A, Almonte A, Sowemimo-Coker G, et al. Disclosure of HIV serostatus and pre-exposure prophylaxis use on internet hookup sites among men who have sex with men. AIDS Behav 2019 Jul 28;23(7):1681-1688 [FREE Full text] [doi: 10.1007/s10461-018-2286-z] [Medline: 30267365]
13. Jozsa K, Kraus A, Korpak AK, Birnholtz J, Moskowitz DA, Macapagal K. "Safe Behind My Screen": adolescent sexual minority males' perceptions of safety and trustworthiness on geosocial and social networking apps. Arch Sex Behav 2021 Oct 28;50(7):2965-2980. [doi: 10.1007/s10508-021-01962-5] [Medline: 34581948]
14. Chu BA, Castellanos ER, Gonzales MM, Gaither TW. Social determinants of self-reported pre-exposure prophylaxis use among a national sample of US men who have sex with men. Sex Transm Infect 2020 Aug 10;96(5):368-374. [doi: 10.1136/sextrans-2019-054177] [Medline: 31601642]
15. Sanchez TH, Zlotorzynska M, Rai M, Baral SD. Characterizing the impact of COVID-19 on men who have sex with men across the United States in April, 2020. AIDS Behav 2020 Jul 29;24(7):2024-2032 [FREE Full text] [doi: 10.1007/s10461-020-02894-2] [Medline: 32350773]
16. Grov C, Zohra F, Westmoreland DA, Mirzayi C, D'Angelo A, Stief M, et al. Sex in the era of COVID-19 in a U.S. national cohort of cisgender men, transgender women, and transgender men who have sex with men: April-May 2020. Arch Sex Behav 2022 Jan 29;51(1):343-354 [FREE Full text] [doi: 10.1007/s10508-021-02121-6] [Medline: 34845579]
17. Chow E, Hocking J, Ong J, Phillips T, Schmidt T, Buchanan A, et al. Brief report: changes in PrEP use, sexual practice, and use of face mask during sex among MSM during the second wave of COVID-19 in Melbourne, Australia. J Acquir Immune Defic Syndr 2021 Feb 01;86(2):153-156 [FREE Full text] [doi: 10.1097/QAI.0000000000002575] [Medline: 33433122]
18. Shilo G, Mor Z. COVID-19 and the changes in the sexual behavior of men who have sex with men: results of an online survey. J Sex Med 2020 Oct;17(10):1827-1834 [FREE Full text] [doi: 10.1016/j.jsxm.2020.07.085] [Medline: 32883631]
19. Starks TJ, Jones SS, Sauermilch D, Benedict M, Adebayo T, Cain D, et al. Evaluating the impact of COVID-19: a cohort comparison study of drug use and risky sexual behavior among sexual minority men in the U.S.A. Drug Alcohol Depend 2020 Nov 01;216:108260 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108260] [Medline: 32890975]
20. Mitchell KM, Dimitrov D, Silhol R, Geidelberg L, Moore M, Liu A, et al. The potential effect of COVID-19-related disruptions on HIV incidence and HIV-related mortality among men who have sex with men in the USA: a modelling study. The Lancet HIV 2021 Apr;8(4):e206-e215. [doi: 10.1016/s2352-3018(21)00022-9]
21. Spinelli M, Hickey M, Glidden D, Nguyen J, Oskarsson J, Havlir D. Viral suppression rates in a safety-net HIV clinic in San Francisco destabilized during COVID-19. Aids 2020;34(15):2328-2331. [doi: 10.1097/qad.0000000000002677]
22. Harkness A, Weinstein ER, Atuluru P, Vidal R, Rodriguez-Diaz CE, Safren SA. "Let's Hook Up When the Pandemic is Over:" Latinx Sexual Minority Men's Sexual Behavior During COVID-19. J Sex Res 2021 Oct 04;58(8):951-957. [doi: 10.1080/00224499.2021.1888064] [Medline: 33661064]

23. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. Lancet Infect Dis 2020 Jun;20(6):678-688. [doi: 10.1016/s1473-3099(20)30162-6]
24. Heath PT, Galiza EP, Baxter DN, Boffito M, Browne D, Burns F, et al. Safety and Efficacy of NVX-CoV2373 Covid-19 Vaccine. N Engl J Med 2021 Sep 23;385(13):1172-1183. [doi: 10.1056/nejmoa2107659]
25. Newport F, Gates G. San Francisco metro area ranks highest in LGBT percentage. Gallup. 2015 Mar 20. URL: https://news.gallup.com/poll/182051/san-francisco-metro-area-ranks-highest-lgbt-percentage.aspx?utm_source=Social%20Issues&utm_medium=newsfeed&utm_campaign=tiles [accessed 2022-08-30]
26. Mantz A, Liebhaber L. Most rural counties in America. Stacker. 2021 Apr 14. URL: https://stacker.com/stories/3271/most-rural-counties-america [accessed 2022-08-30]
27. Census Reporter. URL: https://censusreporter.org/ [accessed 2022-02-07]
28. Blackwell C, Birnholtz J, Abbott C. Seeing and being seen: Co-situation and impression formation using Grindr, a location-aware gay dating app. New Media Soc 2014 Feb 07;17(7):1117-1136. [doi: 10.1177/1461444814521595]
29. Janssen K, Donders A, Harrell F, Vergouwe Y, Chen Q, Grobbee D, et al. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol 2010 Jul;63(7):721-727 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.12.008] [Medline: 20338724]
30. Smith J, Brown S. "No Fats, Femmes, or Blacks:" the role of body types, gender roles and race in condom usage online. In: Gender, Sexuality and Race in the Digital Age. Cham: Springer; Jan 02, 2020:205-222.
31. Babel RA, Wang P, Alessi EJ, Raymond HF, Wei C. Stigma, HIV risk, and access to HIV prevention and treatment services among men who have sex with men (MSM) in the United States: a scoping review. AIDS Behav 2021 Nov 18;25(11):3574-3604 [FREE Full text] [doi: 10.1007/s10461-021-03262-4] [Medline: 33866444]
32. Hubach RD, Currin JM, Giano Z, Meyers HJ, DeBoy KR, Wheeler DL, et al. Experiences of stigma by gay and bisexual men in rural Oklahoma. Health Equity 2019 May;3(1):231-237 [FREE Full text] [doi: 10.1089/heq.2018.0095] [Medline: 31289783]
33. McKenney J, Sullivan PS, Bowles KE, Oraka E, Sanchez TH, DiNenno E. HIV risk behaviors and utilization of prevention services, urban and rural men who have sex with men in the United States: Results from a national online survey. AIDS Behav 2018 Jul 6;22(7):2127-2136. [doi: 10.1007/s10461-017-1912-5] [Medline: 28986669]
34. Owens C, Hubach RD, Williams D, Voorheis E, Lester J, Reece M, et al. Facilitators and barriers of pre-exposure prophylaxis (prep) uptake among rural men who have sex with men living in the Midwestern U.S. Arch Sex Behav 2020 Aug 26;49(6):2179-2191. [doi: 10.1007/s10508-020-01654-6] [Medline: 32219687]
35. Bowen AM, Horvath K, Williams ML. A randomized control trial of Internet-delivered HIV prevention targeting rural MSM. Health Educ Res 2007 Mar 01;22(1):120-127 [FREE Full text] [doi: 10.1093/her/cyl057] [Medline: 16849391]
36. Algarin AB, Ward PJ, Christian WJ, Rudolph AE, Holloway IW, Young AM. Spatial distribution of partner-seeking men who have sex with men using geosocial networking apps: epidemiologic study. J Med Internet Res 2018 May 31;20(5):e173 [FREE Full text] [doi: 10.2196/jmir.9919] [Medline: 29853441]
37. Kesten JM, Dias K, Burns F, Crook P, Howarth A, Mercer CH, et al. Acceptability and potential impact of delivering sexual health promotion information through social media and dating apps to MSM in England: a qualitative study. BMC Public Health 2019 Sep 06;19(1):1236 [FREE Full text] [doi: 10.1186/s12889-019-7558-7] [Medline: 31492120]
38. Bineau L, Lambert D, Truszczynski N, Hansen N, Lauckner C. Dating app use among rural men who have sex with men and its relationship to HIV prevention and risk behaviors: a mixed-methods analysis. RRH 2021 Jun 2. [doi: 10.22605/rrh6556]
39. Parker CM, Garcia J, Philbin MM, Wilson PA, Parker RG, Hirsch JS. Social risk, stigma and space: key concepts for understanding HIV vulnerability among black men who have sex with men in New York City. Cult Health Sex 2017 Mar 23;19(3):323-337 [FREE Full text] [doi: 10.1080/13691058.2016.1216604] [Medline: 27550415]
40. Hatzenbuehler ML, O'Cleirigh C, Mayer KH, Mimiaga MJ, Safren SA. Prospective associations between HIV-related stigma, transmission risk behaviors, and adverse mental health outcomes in men who have sex with men. Ann Behav Med 2011 Oct 30;42(2):227-234 [FREE Full text] [doi: 10.1007/s12160-011-9275-z] [Medline: 21533623]
41. Rogers BG, Coats CS, Adams E, Murphy M, Stewart C, Arnold T, et al. Development of telemedicine infrastructure at an LGBTQ+ clinic to support HIV prevention and care in response to COVID-19, Providence, RI. AIDS Behav 2020 Oct 29;24(10):2743-2747 [FREE Full text] [doi: 10.1007/s10461-020-02895-1] [Medline: 32350772]
42. Hoagland B, Torres TS, Bezerra DR, Geraldo K, Pimenta C, Veloso VG, et al. Telemedicine as a tool for PrEP delivery during the COVID-19 pandemic in a large HIV prevention service in Rio de Janeiro-Brazil. Braz J Infect Dis 2020 Jul;24(4):360-364 [FREE Full text] [doi: 10.1016/j.bjid.2020.05.004] [Medline: 32504552]
43. Hubach RD, O'Neil AM, Stowe M, Hamrick J, Giano Z, Currin JM. Preferred methods of HIV and sexually transmissible infection screening delivery among a rural sample of men who have sex with men. AIDS Patient Care STDS 2020 Nov 01;34(11):470-476. [doi: 10.1089/apc.2020.0170] [Medline: 33147083]
44. Kanny D, Jeffries WL, Chapin-Bardales J, Denning P, Cha S, Finlayson T, National HIV Behavioral Surveillance Study Group. Racial/ethnic disparities in HIV preexposure prophylaxis among men who have sex with men - 23 urban areas, 2017. MMWR Morb Mortal Wkly Rep 2019 Sep 20;68(37):801-806 [FREE Full text] [doi: 10.15585/mmwr.mm6837a2] [Medline: 31536484]

45. Kazi DS, Katz IT, Jha AK. PrEParing to end the HIV epidemic — California's route as a road map for the United States. N Engl J Med 2019 Dec 26;381(26):2489-2491. [doi: 10.1056/nejmp1912293]
46. Arrington-Sanders R, Hailey-Fair K, Wirtz AL, Morgan A, Brooks D, Castillo M, et al. Role of structural marginalization, HIV stigma, and mistrust on HIV prevention and treatment among young Black Latinx men who have sex with men and transgender women: perspectives from youth service providers. AIDS Patient Care STDS 2020 Jan 01;34(1):7-15 [FREE Full text] [doi: 10.1089/apc.2019.0165] [Medline: 31944853]
47. Boone CA, Bowleg L. Structuring sexual pleasure: equitable access to biomedical HIV prevention for Black men who have sex with men. Am J Public Health 2020 Feb;110(2):157-159. [doi: 10.2105/ajph.2019.305503]
48. Sullivan PS, Mena L, Elopre L, Siegler AJ. Implementation strategies to increase PrEP uptake in the south. Curr HIV/AIDS Rep 2019 Aug 8;16(4):259-269 [FREE Full text] [doi: 10.1007/s11904-019-00447-4] [Medline: 31177363]
49. Reyniers T, Rotsaert A, Thunissen E, Buffel V, Masquillier C, Van Landeghem E, et al. Reduced sexual contacts with non-steady partners and less PrEP use among MSM in Belgium during the first weeks of the COVID-19 lockdown: results of an online survey. Sex Transm Infect 2021 Sep 10;97(6):414-419 [FREE Full text] [doi: 10.1136/sextrans-2020-054756] [Medline: 33172917]
50. Chow E, Hocking J, Ong J, Schmidt T, Buchanan A, Rodriguez E, et al. Changing the use of HIV pre-exposure prophylaxis among men who have sex with men during the COVID-19 pandemic in Melbourne, Australia. Open Forum Infect Dis 2020 Jul;7(7):ofaa275 [FREE Full text] [doi: 10.1093/ofid/ofaa275] [Medline: 32704518]
51. Torres TS, Hoagland B, Bezerra DRB, Garner A, Jalil EM, Coelho LE, et al. Impact of COVID-19 pandemic on sexual minority populations in Brazil: an analysis of social/racial disparities in maintaining social distancing and a description of sexual behavior. AIDS Behav 2021 Jan 31;25(1):73-84 [FREE Full text] [doi: 10.1007/s10461-020-02984-1] [Medline: 32737817]
52. Haberer J, Bangsberg D, Baeten J, Curran K, Koechlin F, Amico K, et al. Defining success with HIV pre-exposure prophylaxis: a prevention-effective adherence paradigm. AIDS 2015 Jul 17;29(11):1277-1285 [FREE Full text] [doi: 10.1097/QAD.0000000000000647] [Medline: 26103095]
53. Williamson I, Papaloukas P, Jaspal R, Lond B. 'There's this glorious pill': gay and bisexual men in the English midlands navigate risk responsibility and pre-exposure prophylaxis. Critical Public Health 2018 Jul 17;29(5):560-571. [doi: 10.1080/09581596.2018.1497143]
54. Johnson M. "Not playing that game": an exploratory study of queer men's health information behavior on location-aware dating and sex-seeking mobile applications. University of North Carolina. 2019. URL: https://cdr.lib.unc.edu/concern/masters_papers/1v53k175w [accessed 2022-08-30]
55. McNaghten A, Brewer NT, Hung M, Lu P, Daskalakis D, Abad N, et al. COVID-19 vaccination coverage and vaccine confidence by sexual orientation and gender identity - United States, August 29-October 30, 2021. MMWR Morb Mortal Wkly Rep 2022 Mar 04;71(5):171-176 [FREE Full text] [doi: 10.15585/mmwr.mm7105a3] [Medline: 35113846]
56. Diesel J, Sterrett N, Dasgupta S, Kriss JL, Barry V, Vanden Esschert K, et al. COVID-19 vaccination coverage among adults - United States, December 14, 2020-May 22, 2021. MMWR Morb Mortal Wkly Rep 2021 Jun 25;70(25):922-927 [FREE Full text] [doi: 10.15585/mmwr.mm7025e1] [Medline: 34166331]
57. Reiter PL, Pennell ML, Katz ML. Acceptability of a COVID-19 vaccine among adults in the United States: How many people would get vaccinated? Vaccine 2020 Sep 29;38(42):6500-6507 [FREE Full text] [doi: 10.1016/j.vaccine.2020.08.043] [Medline: 32863069]
58. Szilagyi PG, Thomas K, Shah MD, Vizueta N, Cui Y, Vangala S, et al. National trends in the US public's likelihood of getting a COVID-19 vaccine-April 1 to December 8, 2020. JAMA 2020 Dec 29;325(4):396 [FREE Full text] [doi: 10.1001/jama.2020.26419] [Medline: 33372943]
59. Hui B, Padeniya T, Rebuli N, Gray R, Wood J, Donovan B, et al. A Gonococcal Vaccine Has the Potential to Rapidly Reduce the Incidence of Neisseria gonorrhoeae Infection Among Urban Men Who Have Sex With Men. J Infect Dis 2022 Mar 15;225(6):983-993 [FREE Full text] [doi: 10.1093/infdis/jiab581] [Medline: 34894134]
60. Abara WE, Kirkcaldy RD, Bernstein KT, Zlotorzynska M, Sanchez T. Acceptability of a gonococcal vaccine among sexually active men who have sex with men. Sexual Trans Dis 2021 Jul 13;49(1):76-80. [doi: 10.1097/olq.0000000000001510]
61. Duncan DT, Park SH, Hambrick HR, Dangerfield Ii DT, Goedel WC, Brewer R, et al. Characterizing geosocial-networking app use among young Black men who have sex with men: a multi-city cross-sectional survey in the southern United States. JMIR Mhealth Uhealth 2018 Jun 14;6(6):e10316 [FREE Full text] [doi: 10.2196/10316] [Medline: 29903702]

## Abbreviations

**aOR:** adjusted odds ratio
**CDC:** Centers for Disease Control and Prevention
**GSN:** geosocial networking
**PASS:** Performer Availability Screening Service
**PrEP:** pre-exposure prophylaxis
**STI:** sexually transmitted infection

*Edited by C Basch; submitted 24.03.22; peer-reviewed by M Paton, A Montilha; comments to author 18.04.22; revised version received 24.04.22; accepted 25.04.22; published 22.09.22.*

*Please cite as:*
*Gaither TW, Sigalos JT, Landovitz RJ, Mills JN, Litwin MS, Eleswarapu SV*
*Engagement With HIV and COVID-19 Prevention: Nationwide Cross-sectional Analysis of Users on a Geosocial Networking App*
*J Med Internet Res 2022;24(9):e38244*
*URL: https://www.jmir.org/2022/9/e38244*
*doi:10.2196/38244*
*PMID:36026586*

Original Paper

# The Influence of Provaping “Gatewatchers” on the Dissemination of COVID-19 Misinformation on Twitter: Analysis of Twitter Discourse Regarding Nicotine and the COVID-19 Pandemic

Nathan Silver[1], PhD; Elexis Kierstead[1], MPH; Ganna Kostygina[2], MA, PhD; Hy Tran[2], MS; Jodie Briggs[1], MA, MPP; Sherry Emery[2], MBA, PhD; Barbara Schillo[1], PhD

[1]Schroeder Institute, Truth Initiative, Washington, DC, United States

[2]Social Data Collaboratory, NORC at the University of Chicago, Chicago, IL, United States

**Corresponding Author:**
Nathan Silver, PhD
Schroeder Institute
Truth Initiative
900 G St NW
Washington, DC, 20006
United States
Phone: 1 2024545555
Email: nsilver@truthinitiative.org

## Abstract

**Background:** There is a lot of misinformation about a potential protective role of nicotine against COVID-19 spread on Twitter despite significant evidence to the contrary. We need to examine the role of vape advocates in the dissemination of such information through the lens of the gatewatching framework, which posits that top users can amplify and exert a disproportionate influence over the dissemination of certain content through curating, sharing, or, in the case of Twitter, retweeting it, serving more as a vector for misinformation rather than the source.

**Objective:** This research examines the Twitter discourse at the intersection of COVID-19 and tobacco (1) to identify the extent to which the most outspoken contributors to this conversation self-identify as vaping advocates and (2) to understand how and to what extent these vape advocates serve as gatewatchers through disseminating content about a therapeutic role of tobacco, nicotine, or vaping against COVID-19.

**Methods:** Tweets about tobacco, nicotine, or vaping and COVID-19 (N=1,420,271) posted during the first 9 months of the pandemic (January-September 2020) were identified from within a larger corpus of tobacco-related tweets using validated keyword filters. The top posters (ie, tweeters and retweeters) were identified and characterized, along with the most shared Uniform Resource Locators (URLs), most used hashtags, and the 1000 most retweeted posts. Finally, we examined the role of both top users and vape advocates in retweeting the most retweeted posts about the therapeutic role of nicotine, tobacco, or vaping against COVID-19.

**Results:** Vape advocates comprised between 49.7% (n=81) of top 163 and 88% (n=22) of top 25 users discussing COVID-19 and tobacco on Twitter. Content about the ability of tobacco, nicotine, or vaping to treat or prevent COVID-19 was disseminated broadly, accounting for 22.5% (n=57) of the most shared URLs and 10% (n=107) of the most retweeted tweets. Finally, among top users, retweets comprised an average of 78.6% of the posts from vape advocates compared to 53.1% from others ($z$=3.34, $P$<.001). Vape advocates were also more likely to retweet the top tweeted posts about a therapeutic role of nicotine, with 63% (n=51) of vape advocates retweeting at least 1 post compared to 40.3% (n=29) of other top users ($z$=2.80, $P$=.01).

**Conclusions:** Provaping users dominated discussions of tobacco use during the COVID-19 pandemic on Twitter and were instrumental in disseminating the most retweeted posts about a potential therapeutic role of tobacco use against the virus. Subsequent research is needed to better understand the extent of this influence and how to mitigate the influence of vape advocates over the broader narrative of tobacco regulation on Twitter.

***(J Med Internet Res 2022;24(9):e40331)*** doi:10.2196/40331

**KEYWORDS**

social media; tobacco; COVID-19; nicotine; misinformation; Twitter; information; infodemiology; vaping; therapeutic; influence; environment; harmful; consequences

## Introduction

### Background

During health crises, such as the COVID-19 pandemic, people who are at higher risk of being affected may be more likely to seek health information online [1,2]. However, in the absence of clear, concise, and complete information, digital information channels, such as social media, are often used to help people understand the implications of a health threat [3,4]. Since the use of inhaled nicotine products, such as cigarettes and e-cigarettes, enhances the risk of respiratory illness and places users at greater risk of complications from COVID-19, clear communication about the risks of tobacco use is more important than ever [5-8]. Some have suggested that a widespread respiratory virus, such as COVID-19, could provide an opportunity to amplify public perceptions about the harms of tobacco products [9]. However, conflicting claims about how the virus affected tobacco users quickly emerged [10,11].

A review of early clinical data in Wuhan Province found that smokers were less likely to be admitted to the intensive care unit (ICU) due to COVID-19 complications compared to nonsmokers [10]. Although the study itself did not provide any evidence of a causal relationship between smoking and COVID-19 infection or progression, the authors posited that the anti-inflammatory properties of nicotine might be responsible for the unexpectedly low prevalence of COVID-19–infected smokers in countries with high smoking rates [12,13]. Although no subsequent evidence has been found to support a protective role of nicotine, the notion that smoking, vaping, or nicotine use would prevent COVID-19 circulated, leading researchers to document misinformation about smoking, vaping, and nicotine as being protective against COVID-19 across communication channels, particularly on Twitter [14-16]. Although the spread of problematic information is not unique to Twitter, recent survey data suggest that Twitter users in particular are more likely to recall hearing and believe that nicotine, tobacco, or vaping can prevent COVID-19 [17-20].

The presentation of scientific findings from an early review of clinical records showing fewer smokers than expected among ICU patients as evidence that nicotine prevents COVID-19 is emblematic of the role social media often plays in communications from the scientific establishment to the general public [14]. Such counterintuitive findings are not misinformation in a direct sense, in that they do not present demonstrably false information [21]. Rather, the extrapolation of the study's findings out of context or with overreaching implications exemplifies the sort of claims that are not egregiously false but rather represent unsubstantiated and misleading implications that run counter to the best-available scientific evidence [11,22].

To understand how scientific distortions and misinformation spread on Twitter, it is first important to understand key differences in how traditional news outlets and social media sites disseminate content. Media researchers use variations of the "gates" metaphor to describe how and to what extent elites and other opinion leaders dictate what information passes "through the gates" and on to the masses [23-26]. The first important difference between traditional news media, such as television, print, or even online publications and Twitter, is the elimination of "gatekeepers"—editorial boards and elite decision makers who determine which news receives airtime [27]. However, Twitter's lack of traditional gatekeepers does not mean that the gates controlling the flow of information are left untended. Rather, the most influential users serve as "gatewatchers," who lack absolute control over what passes through the gates but instead heavily influence whether information is channeled into high-traffic areas where it is likely to spread or low-traffic areas where its impact is diluted [28,29]. Thus, rather than a simple 2-step flow of news through a small group of elites to the viewing public, the preferences and ideological lean of a slightly larger group of vocal users dramatically influence what content "trends" in a user-driven marketplace of ideas [30,31].

Existing evidence suggests that the gatewatching framework may be useful to conceptualize how vaping and other tobacco-related information disseminates on Twitter. Previous research has found that social media discourse is predominantly hostile to vaping regulation, prone to exaggerated claims about the health benefits of vaping, and rife with misinformation about vaping and the tobacco industry [18,32-36]. Although the controversial nature of misinformation is often an important factor contributing to its spread, if the loudest and most prolific voices discussing vaping on Twitter are those with a provaping agenda, then such provaping gatewatchers are also a crucial pathway through which misinformation, disinformation, and other problematic or unsubstantiated information spreads on the medium.

Although there is evidence of a provaping bias on Twitter, neither the extent of this bias nor the influence on the volume of pro- versus antivaping content is clear. A recent examination of vaping-related tweets between March and June 2020 found that misinformation about the relationship between COVID-19 and vaping informed chatter that was both pro- as well as antivaping [37]. In a separate study, the same researchers showed that misinformation was endemic to the Twitter discourse about vaping even prior to COVID-19 [36]. Previous research examining the prevalence of the claim that nicotine can prevent COVID-19 found the therapeutic nicotine claim to be prevalent in about 1% of tweets relevant to both the pandemic and tobacco [16]. Building on these findings, we suggest that vape advocates who disproportionately influence the tobacco-related information that trends on Twitter (ie, gatewatchers) were likely instrumental in disseminating content that promoted a therapeutic role of tobacco, nicotine, or vaping against COVID-19.

### This Study

This research examines whether the gatewatching framework can be used to understand how content about a therapeutic benefit of nicotine, tobacco, or vaping disseminated on Twitter. The main premise of the gatewatching framework is that a subset of influential users drives the dissemination of information on Twitter through retweeting content that is consistent with their ideological agenda. Our investigation begins with the assumption that influence on Twitter is concentrated among a small group of top users who produce and disseminate the majority of content. Pew's population-level examination of Twitter behavior supports this assumption [38]. Pew estimates that 97% of tweets are produced by the top 25% of users. Moreover, a high percentage of tweets by these top users are likely to be retweets of other users' original tweets. The influence of provaping gatewatchers would thus be evident in (1) high prevalence of vape advocates among top users, (2) substantial dissemination of ideologically aligned content (eg, tobacco, nicotine, or vaping could prevent COVID-19), and (3) direct evidence of the role of top users and vape advocates in disseminating that content. We thus propose the following research questions (RQs):

- RQ1: How prevalent are vape advocates among the users who produce and disseminate the most content (ie, potential gatewatchers)?
- RQ2: How prevalent was content indicative of provaping advocacy in the broader conversation about tobacco and COVID-19, including (1) top hashtags, (2) top shared Uniform Resource Locators (URLs), and (3) the most retweeted tweets?
- RQ3: What role do top users and vape advocates play in disseminating top content (ie, top retweeted tweets about a therapeutic role of tobacco, nicotine, or vaping against COVID-19)?

## Methods

### Procedure

We began by identifying posts (original tweets and retweets) about COVID-19 from within the entire corpus of tobacco-related tweets posted between January and September 2020. After cleaning and preprocessing the raw data, including removing duplicate posts, we examined the data set at both the post and the user level. At the post level, we conducted a content analysis of the top 1000 retweets during this period to assess the volume of broadly disseminated tweets promoting the preventative nicotine claim compared to the 1% of overall tweets identified in previous research [16]. We then examined the most shared URLs to further quantify how much content promoting a therapeutic benefit of nicotine against COVID-19 was disseminated on Twitter. Next, we examined the user profiles of the most active users (ie, those responsible for the most tweets and retweets) and identified those who posted the original tweets about nicotine preventing COVID-19 among the top 1000 retweets. Finally, we cross-referenced the top user list with those who retweeted original tweets about nicotine preventing COVID-19 to more clearly illuminate the role of these top users in disseminating this content (ie, gatewatching).

### Data Collection

NORC at the University of Chicago maintains a comprehensive archive of tobacco-related Twitter data collected monthly using the Historical Powertrack Application Programming Interface (API) and sorted for relevance by a naive Bayes classifier. Twitter's API allows for targeted searches by keywords that can appear in either the text of the tweet or the metadata. NORC "tapped the firehose" collecting all tweets posted during the study time frame in JSON format and then parsed and merged the data into a data frame at the post level with corresponding variables for username and other relevant metadata. From this broader corpus of tobacco-related tweets posted in the first 9 months of 2020, we developed and validated a keyword filter (Multimedia Appendix 1) to identify tobacco-related posts that were also about COVID-19. We then validated this filter by human-coding a random sample of 2566 original tweets for relevance (precision=0.90, recall=0.89, $F_1$=.89). The text of each tweet was then used to extract important information, including URLs, hashtags, and whether it was an original tweet or a retweet. Counts were then aggregated to provide data frames at the post (tweet or retweet; N=1,420,271), user (N=817,691), and URL (N=54,806) levels and the top 1000 hashtags.

### *Identifying Top Users*

We first sought to identify a smaller group of top users who were clear outliers in terms of the proportion of overall tweets and retweets for which they were responsible. We began with the top 1000 users who posted (both original and retweeted) between 54 and 4897 times each, meaning 0.12% of users were directly (tweeted) or indirectly (retweeted) responsible for 10.93% of all tobacco and COVID-19 content. Among these top 1000, we identified 2 natural inflection points in the data: (1) The median number of posts was 87 (SD 263.10). Only 25 (2.5%) users posted 3 or more SDs from the median number of posts, meaning 2.49% of all activity came from the top 25 users. (2) To expand this list further, we subgrouped the number of tweets per user in bins of 100, with 59% having less than 100 tweets and 83% having less than 200. We thus coded 163 top users who had 200 or more posts and were responsible for 5.59% of all content produced in our data set. Our coded sample of the most influential users averaged 54.15 tweets per month, more than double the threshold for high-volume users set by Pew. We then categorized these top 163 users by identifying at least 1 of 3 criteria in their profiles: (1) explicit mention of vaping or tobacco harm reduction (THR) in the text of the username or profile, (2) a pinned tweet (a tweet that the user chooses to fix to the top of their page) promoting vaping, or (3) at least 3 of their 5 most recent tweets explicitly promoting vaping.

### *Identifying Dissemination of Therapeutic Nicotine Content*

We examined 3 key measures of trending content for dissemination of misinformation related to a potential therapeutic role of nicotine against COVID-19. First, we examined the top trending hashtags during this period. Hashtags are a key means through which social media conversations coalesce around a coherent narrative [39,40]. All hashtags were extracted from the text of the tweets and aggregated. Beginning with the top 1000 hashtags that were used between 50 (#heart) and 87,566 times (#covid19) with a median of 106.50 (SD 3101.53), we identified a natural inflection point in the data wherein only 16 hashtags were used greater than 1 SD from the median, accounting for 49.3% of hashtags used. We then identified hashtags that were explicitly tied to vaping, e-cigarettes, or tobacco harm reduction using keyword stems (eg, vap*, ecig*, thr, and harmreduc*).

Top linked URLs were examined using a similar procedure in finding a natural inflection point in the data to determine the top trending content. URLs were shared a median of 1 time each (SD 24). Of those, 253 URLs were shared greater than 3 SD from the median, comprising 30.9% of all shared URLs. The number of shares for these top URLs ranged from 74 to 2827, with a median of 117.5 (SD 279.73). We then examined these top URLs to determine whether they were linking content that promoted the ability of nicotine, tobacco, or vaping to prevent or treat COVID-19.

Finally, we conducted a content analysis of the top 1000 retweets to characterize the presence of original tweets about a potential therapeutic benefit of nicotine, tobacco, or vaping against COVID-19 in the most broadly disseminated part of the broader conversation about tobacco and COVID-19 on Twitter. We used a grounded theory approach [41,42]. We reviewed the top 1000 retweets while noting 6 relevant themes, with the primary theme of interest being the potential therapeutic role of nicotine or tobacco against COVID-19. Consistent with the convention for content analyses, a random subsample of at least 10% was withheld to establish reliability [43]. In this study, 2 independent coders dual coded a random subsample of 300 (30%) retweets to establish reliability in identifying tweets about personal responsibility (κ=0.95), social justice (κ=0.83), discounting COVID-19 severity compared to tobacco (κ=1), government criticism (κ=0.92), mask efficacy (κ=0.8), and, the topic of interest, the protective role of nicotine (κ=1). After establishing reliability, the remaining 700 (70%) retweets were divided evenly among the coders. The user profiles of those whose retweets promoted a therapeutic role of nicotine against COVID-19 were then coded to identify vaping advocates using the same methodology as was used for coding the top users. Finally, our RQ about gatewatching was then examined by identifying whether the top users retweeted the top retweeted content about a potential therapeutic role of nicotine or tobacco.

### *Bot Detection*

The role of automated (bot) accounts on Twitter has been a recent area of concern [44]. One report suggested that as many as half of all tweets about vaping may come from bots [44]. Although a bot programmed to promote vaping content serves functionally the same purpose as a human gatewatcher who promotes vaping content, differentiating between bots and human accounts is important, as regulatory bodies and health communicators are likely to approach these sources of problematic information in different ways [45]. We first used the machine learning classifier *Botometer* to estimate the likelihood that user accounts were bots based on a series of indicators of "botlike" behaviors identified by the tool's creators, providing a score between 0 and 5, with 5 being the most likely to be a bot [46]. However, this tool has been shown to have significant limitations in misclassifying both bot and human accounts [47]. As a result, we reported additional indicators that may be indicative of bot activity, including whether the account is verified by Twitter and whether the account has since been removed or made private. It is noteworthy that the top user overall had a bot score of 3.6 and only posted original tweets. However, this user was not a vape advocate.

### Ethical Considerations

Human subjects were not involved in this study. Data were collected from public social media sites. Account names were excluded for anonymity in publication.

## Results

### Vape Advocates and Vaping Hashtags

Vape advocates were highly prevalent across the top users, representing 81 (49.7%) of the top 163 users and 22 (88%) of the top 25 users. These top 163 users posted a median of 317 times each (SD 536.56) and retweeted (median 234, SD 373.80) far more often than they posted original tweets (median 58, SD 433.98). On average, retweets comprised a higher percentage of posts for vape advocates (78.6%) than for others (53.1%; $z$=3.34, $P$<.001). The prevalence of bots among top users appeared limited. Although only 3 accounts were verified by Twitter and 20 were either removed or private, the average Botometer score was low (mean 1.59, SD 1.37), with only 19% (n=31) of users having above the scale's midpoint of 2.5. The average score for vape advocates was 1.26 but 2.03 for all other accounts, providing little evidence that bots are driving vape advocacy in our data set.

Vaping hashtags (n=63) were used a total of 43,223 times, accounting for 9.4% of the hashtags used in our data set, including 3 of the top 16 most used overall. Table 1 provides the top 16 overall hashtags in the data set as well as the top 16 vaping hashtags, accounting for 85.7% of the vaping hashtags used. Most noteworthy is the use of #wevapewevote among the top overall hashtags as well as 5 other explicitly provaping hashtags, with over 1000 uses each.

Of the 253 top shared URLs, 57 (22.5%) promoted content about a potential therapeutic role of nicotine, tobacco, or vaping in treating or preventing COVID-19. These URLs were shared 16,244 times. Table 2 provides descriptions of the top 29 shared URLs identified via an inflection point in the data at more than 2 SD from the median number of shares, accounting for 12.4% of all shared URLs. Among these top 29 URLs, 12 URLs linked articles promoting the potential therapeutic value of nicotine, vaping, or tobacco against COVID-19, accounting for 41.4% of shares among this top content. It is noteworthy that 2 (17%) of these 12 were articles explicitly debunking the claim made by a media personality that vaping bleach could cure COVID-19, while the other 10 (83%) focused on either a lower infection rate of COVID-19 for smokers (n=9, 90%) or a tobacco-based vaccine (n=1, 10%).

The top 1000 retweeted posts were shared a total of 578,763 times, ranging between 105 and 117,662, with a median of 193 (SD 3956.82). Table 3 provides the 6 coded categories, example tweets, and the percentage of retweets in each category. The therapeutic potential of nicotine or tobacco was the fourth-most commonly discussed topic. Of the 107 retweeted posts addressing the protective role of nicotine, including smoking, vaping, or tobacco in general, 5 (4.7%) sought to counter this notion and accounted for 1304 (0.2%) retweets. Closer examination also revealed that 4 of these retweets (3.7%) concerned addressing a conservative talk show host who told a call-in listener that they could vape bleach to protect themselves from COVID-19. After removing these 9 (8.4%) tweets, we focused on 98 (91.6%) of the top retweets explicitly endorsing or promoting the idea that nicotine, whether through patches, smoking, or vaping, could prevent COVID-19. Such content was retweeted 21,782 times, garnering 3.8% of retweets in our sample (median 160, SD 194.25). Moreover, these tweets also used the hashtags #saysscience (n=17, 17%) and #sciencesurprises (n=12, 12%), which were used across 2129 and 1544 retweets, respectively.

A total of 74 unique users produced the 98 top retweets about a therapeutic role of nicotine. Of these, 30 (40.5%) were verified, while 16 (21.6%) were official news accounts. In fact, retweets by verified accounts garnered 74.5% of retweets, while news accounts (all but 2 of which were verified) garnered 46.2% of retweets. There were only 2 (2.7%) vape advocates among these 74 users. Finally, bots were limited among this group as well, with an average Botometer score of 1.57 (SD 1.28). Notably, this value is likely inflated as the verified news sources tended to be misclassified as bots, with an average score of 3.19.

The top 163 users retweeted the top posts about nicotine preventing COVID-19 338 times (median 1, SD 2.78). Among the top 163 users, 91 (55.8%) retweeted at least 1 of the top posts, with 17 (68%) of the top 25 retweeting at least 1 post (median 4, SD 4.51). A significantly higher percentage of vape advocates (63%) retweeted such posts compared to other top users (40.3%; z=2.80, *P*=.01). In total, 38.2% of top posts were retweeted at least once by top users, with original posts by the lead author of the study showing a lower-than-expected number of smokers in the ICU with COVID-19 garnering 38.5% of retweets by top users. Table 4 provides the deidentified text of the top retweets promoting such content retweeted by top users.

**Table 1.** Top hashtags for COVID-19– and nicotine-related discussions on Twitter.

| Top 16 overall hashtags[a] (N=87,566) | | Top 16 vaping hashtags[b] (N=43,223) | |
|---|---|---|---|
| Hashtags | Uses, n (%) | Hashtags | Uses, n (%) |
| covid19 | 87,566 (100) | vaping[c] | 15,567 (36.0) |
| coronavirus | 31,608 (36.1) | vape[c] | 6306 (14.6) |
| vaping | 15,567 (17.8) | wevapewevote | 3601 (8.3) |
| nomeat_nocoronavirus | 11,495 (13.1) | vapingsaveslives | 2419 (5.6) |
| tobacco | 10,961 (12.5) | ecigs[c] | 1495 (3.5) |
| covid-19 | 9181 (10.5) | vapefam | 1283 (3.0) |
| smoking | 8328 (9.5) | harmreduction | 1137 (2.6) |
| covid | 8193 (9.4) | vapers | 1137 (2.6) |
| lockdownsa | 7685 (8.8) | ecigarettes[c] | 871 (2.0) |
| covid_19 | 7256 (8.3) | ecig[c] | 766 (1.8) |
| stayhome | 6705 (7.7) | vapelife | 734 (1.7) |
| vape | 6306 (7.2) | vapes | 502 (1.2) |
| quitforcovid | 5156 (5.9) | tobaccoharmreduction | 424 (1.0) |
| lockdown | 4021 (4.6) | vapeon | 414 (1.0) |
| wevapewevote | 3601 (4.1) | vapecommunity | 390 (0.9) |
| indiafightscorona | 3472 (4.0) | vapenation | 387 (0.9) |

[a] The top 16 hashtags accounted for 49.3% of all hashtags used.

[b] The top 16 vaping hashtags accounted for 85.71% of all provaping hashtags used and 9.4% of all hashtags used.

[c] These hashtags were not necessarily provaping and were sometimes used in antivaping posts as well.

**Table 2.** Top 29[a] URLs[b] shared in COVID-19– and nicotine-related discussions on Twitter.

| URL description | Shares (N=21,100), n (%) |
|---|---|
| News24.com:lockdown dlamini zuma pushes for tobacco alcohol ban to continue until level 1 | 2827 (13.4) |
| The economist: smokers seem less likely than non smokers to fall ill with covid 19 | 2359 (11.2) |
| News24.com: coronavirus all the latest news about covid 19 in south africa and the world | 1854 (8.8) |
| Raw story: conservative radio host agrees with caller that vaping bleach might cure covid 19 youre not crazy | 878 (4.2) |
| The Guardian: French study suggests smokers at lower risk of getting coronavirus | 796 (3.8) |
| News24.com: breaking ramaphosa told to lift cigarette alcohol ban and move to level 2 lockdown sources | 694 (3.3) |
| rFi.fr: french researchers suggest nicotine could protect against covid 19 | 677 (3.2) |
| France24.com: france testing whether nicotine could prevent coronavirus 1 | 667 (3.2) |
| CNN: coronavirus quitting smoking wellness | 657 (3.1) |
| News 24: not selling booze and tobacco during lockdown harmful to addicts | 649 (3.1) |
| ewn.co.za: sa economy loses r1 5 billion due to alcohol cigarette sale ban | 631 (3.0) |
| Telegraph: smokers four times less likely contract covid 19 prompting nicotine/ | 620 (2.9) |
| News24.com: coronavirus all the latest news about covid 19 in south africa and the world | 584 (2.8) |
| Livemint.com: cigarette can keep coronavirus away researchers test if nicotine could prevent covid 19 | 557 (2.6) |
| Media Matters: Sean Hannity suggests vaping prevents people getting coronavirus | 491 (2.3) |
| Money Control: a cigarette a day can keep coronavirus away french researchers test if nicotine can prevent covid | 488 (2.3) |
| Bloomberg: coronavirus vaccine race gets unlikely partner big tobacco | 472 (2.2) |
| W24.co.za: How lockdown saved my life woman shares how she finally quit smoking after 20 years | 468 (2.2) |
| Scientific American: Smoking or vaping may increase the risk of a severe coronavirus infection | 461 (2.2) |
| CNN: coronavirus quitting smoking wellness | 461 (2.2) |
| The Guardian: french study suggests smokers at lower risk of getting coronavirus | 457 (2.2) |
| NDTV: coronavirus drug news france testing if nicotine prevents coronavirus from attaching to cells | 444 (2.1) |
| Nature: Factors associated with COVID-19-related death using OpenSAFELY | 435 (2.1) |
| The Guardian: Politics public covid 19 tobacco johnson | 431 (2.0) |
| CNN: coronavirus quitting smoking wellness | 423 (2.0) |
| Zerohedge.com: Did china steal coronavirus from canada and weaponize it | 409 (1.9) |
| Buzzfeed: smoking doesnt kill and other great old op eds from mike pence | 408 (1.9) |
| Medium: how i killed the smoke monster and quit smoking like a queen | 402 (1.9) |
| Todayistheday.co.uk/ (Resources for smoking cessation) | 400 (1.9) |

[a] These 29 web articles accounted for 12.4% of shares among a total of 54,806 different URLs that were shared a combined total of 170,496 times.

[b] URL: Uniform Resource Locator.

**Table 3.** Content analysis of the top retweets (N=1000)[a] about COVID-19 and nicotine.

| Coded category | Example | Top retweets, n (%) |
|---|---|---|
| Government criticism | MIKE PENCE: - His budget cuts in Indiana led to the spread of HIV there - Wrote articles about how smoking does NOT cause cancer - Calls global warming a myth - Was put in charge, by trump, of the Coronavirus response UNBELIEVABLE. #CoronaVirusUpdates | 293 (29.3) |
| Personal responsibility | As COVID-19 attacks the lungs one of the most important things you can do is to quit smoking and vaping. I'm in day 3. Care to join? | 149 (14.9) |
| Mask efficacy | Ok I recorded the recording because I know they will remove it... why?? Vape smoke is 2.5 microns... Covid is between 0.15-0.25 microns. Masks don't do shit. | 149 (14.9) |
| Protective role of nicotine | Nicotine could protect people from contracting the coronavirus, according to new research in France, where further trials are planned to test whether the substance could be used to prevent or treat the deadly illness | 107 (10.7) |
| Discounting the pandemic's impact | If they're going to report every Coronavirus death, I think they should have to report every: Flu Death Car Accident Death Smoking-related Death Alcohol Related Death . . . You get the point. ENOUGH WITH THE FEAR MONGERING | 50 (5.0) |
| Social justice | imagine if the surgeon general announced a plan to bolster access to masks, testing, &; neighborhood health hubs for Black & Latinx people instead of telling us to not smoke &; drink to protect big momma'n'em. how do you blame people for being imperfect victims of a pandemic? | 49 (4.9) |

[a] Of the 1000 posts, 2 independent coders double-coded 300 (30%) posts to establish reliability, after which the remaining 700 (70%) were divided evenly between the 2 coders.

**Table 4.** Top 10 posts[a] about nicotine preventing COVID-19 retweeted by top users.

| Tweet text | Top users who retweeted, n (%) |
|---|---|
| NYC Mayor said smoking and vaping increases coronavirus risk. In 1099 cases from China, only 12.6% were smokers (we would expect much higher). ZERO data on e-cigs. So, still too early to say. People with ZERO public health knowledge should SHUT UP. | 39 (23.9) |
| Finally, the study is out "Systematic review of the prevalence of current smoking among hospitalized COVID-19 patients in China could nicotine be a therapeutic option?" Very low prevalence of smoking among hospitalized COVID-19 patients in China. | 38 (23.3) |
| Moderate and heavy smokers were 50-60% less likely to be tested positive for COVID-19 and 80-90% less likely to be admitted to the ICU... Remember my hypothesis about the potentially protective effects of nicotine since early April? | 31 (19.0) |
| So few people hospitalized with the coronavirus appear to be smokers. I spoke to the scientists, tobacco experts, and policymakers who are trying to see if nicotine *might* have something to do with it. | 23 (14.1) |
| Dramatic UNDER-representation of smokers among COVID-19 patients in France. 80% reduced standardized (for age and sex) incidence ratio!! Strongly supports my hypothesis about the protective effects of nicotine which i made 1 month ago (soon to be published). | 22 (13.5) |
| On January 22 at the beginning of this year I had a suspicion about the protective effect of nicotine on the coronavirus. | 20 (12.3) |
| Official French data on #tobacco smoking; #covid19 replicate the picture in China, Germany; USA A remarkable low rate of smokers are hospitalised w/ coronavirus compared to smoking prevalence (France 23%). | 18 (11.0) |
| The government has admitted "smoking populations were less likely to be infected" with the coronavirus and develop Covid-19. | 12 (7.4) |
| The prohibition of cigarettes does nothing to control or limit the Covid 19 epidemic. On the contrary smokers are significantly less likely to require hospitalization if they do become infected. NDZ is pursuing a very personal and subjective campaign. | 12 (7.4) |
| "There is zero evidence that smoking will propagate or increase transmission of COVID-19."—Dr Konstantinos Farsalinos, Cardiologist and anti-smoking researcher | 11 (6.8) |

[a] These top 10 posts accounted for 69.8% of retweets of such content by top users; 29 other posts were retweeted by between 1 and 9 top users.

# Discussion

## Principal Findings

This research demonstrates the utility of the gatewatching framework for examining the dissemination of problematic information on Twitter. More than half of our sample of top users and 22 of the 25 most prolific users producing and disseminating content about COVID-19 and tobacco in the first 9 months of the pandemic were provaping "harm reduction" advocates. Moreover, more than 3 of 4 posts by these top-using vape advocates were retweets compared to just over half for non-vape advocates, further demonstrating the key role of these users as disseminators of content—gatewatchers.

Building on previous research both identifying and quantifying the extent of a specific piece of misinformation that nicotine can prevent COVID-19, we showed the disproportionately broad reach of this claim across the most retweeted content during this period [16,37,48]. Even in May 2020, when the original study by Farsalinos et al [10] was published, the preponderance of scientific evidence, including multiple meta-analyses, still opposed the notion that nicotine, and especially smoking, would protect people from COVID-19. Still, in our sample of the top 1000 retweets, propagation of this claim was more than 20 times more common than the 5 tweets trying to debunk the claim and was retweeted nearly 17 times as often. Among the top shared URLs, articles promoting a potential therapeutic role of nicotine or tobacco accounted for nearly 1 in 3 shares. Explicitly provaping hashtags, such as #wevapewevote, #vapingsaveslives, and #vapefam, were abundant, indicating a significant representation of a provaping perspective across the broader COVID-19– and tobacco-related conversation.

Finally, this study provides compelling evidence that the top users, particularly vape advocates, were instrumental in disseminating the idea that nicotine can prevent COVID-19. Vape advocates were significantly more likely to retweet top tweets about a potential therapeutic role of nicotine. More than half of the top users retweeted at least 1 of the most retweeted tweets on the topic, while more than 1 in 3 of the most retweeted tweets was retweeted by at least 1 top user. These findings have implications for both tobacco control and the process of disseminating information on Twitter.

## Implications for Tobacco Control

The most important implication for tobacco control is that the dissemination of tobacco content on Twitter is heavily influenced by vape advocates. The extent to which COVID-19 served as further motivation for smokers or vapers trying to quit is uncertain. However, research examining this question has found mixed results at best [49-51]. Although the use of addictive substances during a pandemic is explained by far more variables than misinformation on social media, our findings suggest that the provaping ideological bias of the most prominent voices on Twitter engaging in the tobacco control conversation may help explain why misinformation promoting the protective role of nicotine was disseminated so broadly. Furthermore, previous research examining temporal trends in the tobacco sentiment on Twitter noted an increase in antitobacco sentiment in March 2020 at the beginning of the pandemic in the United States, followed by a rise in positive tobacco sentiment corresponding with the release of the preprint of the study showing fewer smokers than expected in Wuhan ICUs [52]. Our study shows that a likely driver of positive sentiment—that tobacco, nicotine, or vaping has therapeutic value against COVID-19—was spread frequently by influential vape advocates on the platform.

This research highlights the growing challenge of addressing scientific distortions that while not themselves misinformation can nonetheless drive false beliefs. There is no reason to believe that the study finding fewer smokers than expected was falsified. In fact, this “smoker’s paradox” drove significant research interest and calls to pre-register hypotheses toward the goal of rigorously investigating the effects of nicotine on COVID-19 [53]. A substantial body of literature has provided strong evidence that smoking during a respiratory pandemic increases the risk of severe illness and death [5-8]. Moreover, more detailed investigations of specific hypotheses surrounding a therapeutic effect of nicotine have revealed the opposite, as nicotine appears to aid the replication of SARS-CoV-2 rather than impede it [54]. Our study does not address this complicated body of literature. Rather, we show how an opportunistic overinterpretation of the findings of such a study can disseminate on Twitter through influential users for whom such findings support a broader narrative. The broader implication of these findings is that the dissemination of information about tobacco control on Twitter is subject to the interpretation of users who both strongly influence what information is disseminated and whose stated purpose on the platform is to oppose tobacco regulation.

## Implications for Understanding Misinformation Dissemination

It is important to note that our findings do not contradict previous work examining the prevalence of misinformation on the protective role of nicotine but rather add context that helps to characterize the process through which misinformation spreads on Twitter. Kavaluru et al [16] identified that the protective role of nicotine constitutes about 1% of the overall content, while Sidani et al [37] identified a variety of different misinformation claims that arose on Twitter about vaping products. Both studies provide an overview of the overall “firehose” of information. We sought to understand how a small percentage of the information from this firehose got diverted into the smaller and more influential pool of trending retweets. Although we do not discount the fundamental virality of controversy that previous research suggests can drive the dissemination of misinformation, we highlight the important utility of the “gatewatcher” metaphor in describing how misinformation disseminates on Twitter [55].

Opinion leaders on Twitter do not have control over the content posted on the platform. However, they have outsized influence over the dissemination of certain perspectives over others. Although more research is needed, we contribute strong evidence that the ideological lean of the most prolific tweeters on a given subject (provape users discussing COVID-19 and nicotine) directly influenced the spread of problematic information (that nicotine could prevent COVID-19) through retweeting much of the most broadly disseminated posts. These findings are reminiscent of previous research showing the majority of disinformation in another context, antivaccination, emanated from only 12 users [56]. However, in contrast to the “Disinformation Dozen,” provaping gatewatchers on Twitter do not produce and disseminate overtly false information. Rather, they serve as mediators between the scientific community and the broader, Twitter-using public, and privilege scientific findings that support a provaping narrative, while dismissing, ignoring, and countering a preponderance of evidence that does not.

### Implications for Practitioners

There are 2 useful implications of this research for practitioners. The first is that a small component of the overall conversation can have an outsized influence. Ultimately, most of the conversation about COVID-19 and nicotine was about COVID-19 and only tangentially mentioned nicotine or tobacco products. Although #wevapewevote was among the top trending hashtags, #nomeatnocoronavirus was over 3 times more prevalent. Moreover, 60% of our sample of top retweets involved criticizing the government, complaining about masks, and extolling the virtues of taking care of one's body during the pandemic.

Social media's unprecedented democratization of the fourth estate unfortunately nests genuine attempts to engage constructive public discourse alongside incoherent, misinformed, and often bad-faith commentary [57]. Trending hashtags and overall prevalence tell an important part of the story with regard to what and how information spreads on social media [58]. That controversy is interesting and that anybody can post anything with limited oversight are both established and intuitive reasons why misinformation is endemic to social media [59]. However, addressing such a problem requires a closer examination of the vectors through which some tweets spread while others do not. Misinformation about nicotine and COVID-19 does not comprise a majority or even a plurality of content about using nicotine products during the pandemic. However, misinformation about nicotine preventing COVID-19 circulated broadly most likely because it was consistent with the ideological agenda of the opinion leaders, gatewatchers, and most prolific tweeters on the subject.

The broader implication of this process of dissemination is that the proverbial deck is stacked against effective public health communication on Twitter. In the context of nicotine's potential role in preventing COVID-19, Twitter undoubtedly amplified bad information when good information was available. The observable provaping bias of the most outspoken users discussing COVID-19 and nicotine inevitably meant that even attempts to debunk such information on the platform did not receive nearly the same amount of traffic. The most important implication is that this bias is likely to result in the continued prominence of the benefits of vaping, while underrepresenting and downplaying the harms.

### Limitations and Future Directions

The most important limitation of this research is related to the scope of our findings. We examined the influence of gatewatchers on 1 of many social media platforms, in the specific context of tobacco control and COVID-19 during the onset of the pandemic. The dissemination of misinformation on social media is likely to vary between different platforms, different contexts, and potentially within the broader context of tobacco control. Although our findings have generalizable implications, more research is needed to fully understand the interplay between platform, context, and specific kinds of mis- and disinformation, including distortions of scientific consensus.

A second limitation concerns the conclusiveness of our findings with regard to the central premise of the gatewatching framework—that top users were directly responsible for the broad dissemination of the most retweeted content. We provide conclusive evidence that the top users discussing COVID-19 in the context of tobacco, most of whom were vape advocates, retweeted many of the most broadly disseminated retweets about nicotine as potentially therapeutic against COVID-19. However, network approaches are needed in subsequent research to identify whether it is indeed the retweets by these vape advocates that catalyze the broad dissemination of this and other content. Moreover, whether the influence of vape advocates extends beyond the intersection of COVID-19 and tobacco into the broader discussion of tobacco regulation on Twitter and social media is an important topic for subsequent research.

Additionally, we analyzed a small fraction of a data set comprising nearly 1.5 million posts and hundreds of thousands of users. However, research examining misinformation networks online suggests that the best way to reduce misinformation is to identify and penalize the central nodes—the opinion leaders and gatewatchers who drive the virality of some information over others [60]. Although the sheer quantity of content produced is an admittedly blunt instrument for assessing influence, we show that the primary function of these top users, and to an even greater extent the vape advocates among these top users, is to amplify (ie, retweet) some original tweets over others. Subsequent research should adopt more sophisticated measures to assess sustained influence over tobacco regulatory discourse, as a whole, toward which intervention is most likely to be effective.

Finally, although our use of the API and extensive filtering from the broader tobacco-related discussion on Twitter is a strength of this study in providing a near census of relevant content, limitations related to the collection of data from private and removed accounts mean that there is inevitably some content we missed. We note that we were able to capture content from users whose accounts were made private after we collected data.

## Conclusion

The COVID-19 pandemic offered a potential opportunity to highlight the importance of respiratory health by underscoring the negative long-term consequences of inhaled nicotine product use. However, the ability of provaping opinion leaders or gatewatchers on Twitter to steer the narrative and promote misinformation about nicotine as protective against COVID-19 likely played a role in dampening any positive effect of the pandemic on tobacco use. Although anyone can post on Twitter, the makeup of Twitter's tobacco and COVID-19 opinion leaders suggests that content about the dangers of tobacco and vaping does not spread with the same virality as messages that support the proliferation of vaping.

## Acknowledgments

We would like to acknowledge the efforts of Zahra Ehtesham and Lan Margosis who content-coded the top retweets with excellent reliability.

---

## Conflicts of Interest

None declared.

---

Multimedia Appendix 1
Top 30 search terms identifying COVID-19 tweets.
[DOCX File , 14 KB - jmir_v24i9e40331_app1.docx ]

---

## References

1. Roberts M, Ehde D, Herring T, Alschuler K. Public health adherence and information-seeking for people with chronic conditions during the early phase of the COVID-19 pandemic. PM R 2021 Nov;13(11):1249-1260 [FREE Full text] [doi: 10.1002/pmrj.12668] [Medline: 34218517]
2. Vanderpool RC, Huang GC, Mollica M, Gutierrez AI, Maynard CD. Cancer information-seeking in an age of COVID-19: findings from the National Cancer Institute's Cancer Information Service. Health Commun 2021 Jan 22;36(1):89-97. [doi: 10.1080/10410236.2020.1847449] [Medline: 33225770]
3. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. J Med Internet Res 2021 Jun 11;23(6):e29802 [FREE Full text] [doi: 10.2196/29802] [Medline: 34043526]
4. Shearer E. More Than Eight-in-Ten Americans Get News from Digital Devices. 2021. URL: https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/ [accessed 2022-09-10]
5. Alqahtani JS, Oyelade T, Aldhahir AM, Alghamdi SM, Almehmadi M, Alqahtani AS, et al. Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: a rapid systematic review and meta-analysis. PLoS One 2020 May 11;15(5):e0233147 [FREE Full text] [doi: 10.1371/journal.pone.0233147] [Medline: 32392262]
6. Gupta A, Nethan S, Mehrotra R. Tobacco use as a well-recognized cause of severe COVID-19 manifestations. Respir Med 2021 Jan;176:106233 [FREE Full text] [doi: 10.1016/j.rmed.2020.106233] [Medline: 33253975]
7. Patanavanich R, Glantz S. Smoking is associated with COVID-19 progression: a meta-analysis. Nicotine Tob Res 2020;22(9):1653-1656.
8. Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, et al. The impact of COPD and smoking history on the severity of COVID-19: a systemic review and meta-analysis. J Med Virol 2020 Oct;92(10):1915-1921 [FREE Full text] [doi: 10.1002/jmv.25889] [Medline: 32293753]
9. Popova L. Carpe covid: using COVID-19 to communicate about harms of tobacco products. Tob Control 2022 May 13;31(3):397-398. [doi: 10.1136/tobaccocontrol-2020-056276] [Medline: 33188149]
10. Farsalinos K, Barbouni A, Niaura R. Smoking, vaping and hospitalization for COVID-19. Qeios Preprint posted online April 4, 2020. [doi: 10.32388/Z69O8A.13]
11. Majmundar A, Allem J, Cruz TB, Unger JB. Public health concerns and unsubstantiated claims at the intersection of vaping and COVID-19. Nicotine Tob Res 2020 Aug 24;22(9):1667-1668 [FREE Full text] [doi: 10.1093/ntr/ntaa064] [Medline: 32285129]
12. Farsalinos K, Niaura R, Le Houezec J, Barbouni A, Tsatsakis A, Kouretas D, et al. Editorial: nicotine and SARS-CoV-2: COVID-19 may be a disease of the nicotinic cholinergic system. Toxicol Rep 2020;7:658-663 [FREE Full text] [doi: 10.1016/j.toxrep.2020.04.012] [Medline: 32355638]
13. ClinicalTrials. Efficacy of Nicotine in Preventing COVID-19 Infection (NICOVID-PREV). 2021. URL: https://clinicaltrials.gov/ct2/show/NCT04583410 [accessed 2022-09-10]
14. Bandara NA, Herath J, Mehrnoush V. Addressing e-cigarette health claims made on social media amidst the COVID-19 pandemic. World J Pediatr 2021 Feb 21;17(1):3-5 [FREE Full text] [doi: 10.1007/s12519-021-00410-7] [Medline: 33475917]
15. Majmundar A, Allem J, Unger JB, Cruz TB. Vaping and COVID-19: insights for public health and clinical care from Twitter. Int J Environ Res Public Health 2021 Oct 26;18(21):11231 [FREE Full text] [doi: 10.3390/ijerph182111231] [Medline: 34769751]
16. Kavuluru R, Noh J, Rose S. Twitter discourse on nicotine as potential prophylactic or therapeutic for COVID-19. medRxiv Preprint posted online September 18, 2021 [FREE Full text] [doi: 10.1101/2021.01.05.21249284] [Medline: 33442710]
17. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
18. Tan AS, Bigman CA. Misinformation about commercial tobacco products on social media—implications and research opportunities for reducing tobacco-related health disparities. Am J Public Health 2020 Oct;110(S3):S281-S283. [doi: 10.2105/ajph.2020.305910]

19. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
20. Silver NA, Kierstead EC, Briggs J, Schillo B. Charming e-cigarette users with distorted science: a survey examining social media platform use, nicotine-related misinformation and attitudes towards the tobacco industry. BMJ Open 2022 Jun 01;12(6):e057027 [FREE Full text] [doi: 10.1136/bmjopen-2021-057027] [Medline: 35649587]
21. Vraga EK, Bode L. Defining misinformation and understanding its bounded nature: using expertise and evidence for describing misinformation. Political Commun 2020 Feb 06;37(1):136-144. [doi: 10.1080/10584609.2020.1716500]
22. Southwell BG, Brennen JSB, Paquin R, Boudewyns V, Zeng J. Defining and measuring scientific misinformation. Ann Am Acad Pol Soc Sci 2022 May 05;700(1):98-111. [doi: 10.1177/00027162221084709]
23. Bruns A. Gatekeeping, gatewatching, real-time feedback: new challenges for journalism. BJR 2011 Dec 09;7(2):117-136. [doi: 10.25200/bjr.v7n2.2011.355]
24. Pearson GDH, Kosicki GM. How way-finding is challenging gatekeeping in the digital age. Journal Stud 2016 Jan 12;18(9):1087-1105. [doi: 10.1080/1461670x.2015.1123112]
25. Shoemaker PJ, Timothy V. Gatekeeping Theory. New York, NY: Routledge; 2009.
26. Vos T, Heinderyckx F. Gatekeeping in Transition. New York, NY: Taylor & Francis; 2015.
27. Bro P, Wallberg F. Digital gatekeeping. Digit Journal 2014 Apr 14;2(3):446-454. [doi: 10.1080/21670811.2014.895507]
28. Bruns A. Gatewatching and news curation: journalism, social media, and the public sphere. In: Digital Formations, Volume 113. Bern, Switzerland: Peter Lang; 2018.
29. Bruns A, Highfield T. Chapter 18: from news blogs to news on Twitter: gatewatching and collaborative news curation. In: Handbook of Digital Politics. Cheltenham, UK: Edward Elgar; 2015.
30. Krittanawong C, Narasimhan B, Virk H, Narasimhan H, Hahn J, Wang Z. Misinformation dissemination in Twitter in the COVID-19 era. Am J Med 2020;133(12):1367-1369. [doi: 10.1016/j.amjmed.2020.07.012]
31. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? BMJ 2014 Oct 14;349:g6178. [doi: 10.1136/bmj.g6178] [Medline: 25315514]
32. Albarracin D, Shavitt S. Attitudes and attitude change. Annu Rev Psychol 2018 Jan 04;69(1):299-327. [doi: 10.1146/annurev-psych-122216-011911] [Medline: 28841390]
33. Kwon M, Park E. Perceptions and sentiments about electronic cigarettes on social media platformsystematic review. JMIR Public Health Surveill 2020;6(1):e13673. [doi: 10.2196/13673]
34. Lazard A, Saffer A, Wilcox G, Chung A, Mackert M, Bernhardt J. E-cigarette social media messages: a text mining analysis of marketing and consumer conversations on Twitter. JMIR Public Health Surveill 2016;2(2):e6551. [doi: 10.2196/publichealth.6551]
35. McCausland K, Maycock B, Leaver T, Wolf K, Freeman B, Jancey J. E-cigarette advocates on Twitter: content analysis of vaping-related tweets. JMIR Public Health Surveill 2020 Oct 14;6(4):e17543 [FREE Full text] [doi: 10.2196/17543] [Medline: 33052130]
36. Sidani JE, Hoffman BL, Colditz JB, Melcher E, Taneja SB, Shensa A, et al. E-cigarette-related nicotine misinformation on social media. Subst Use Misuse 2022 Jan 22;57(4):588-594. [doi: 10.1080/10826084.2022.2026963] [Medline: 35068338]
37. Sidani JE, Hoffman B, Colditz JB, Wolynn R, Hsiao L, Chu K, et al. Discussions and misinformation about electronic nicotine delivery systems and COVID-19: qualitative analysis of Twitter content. JMIR Form Res 2022 Apr 13;6(4):e26335 [FREE Full text] [doi: 10.2196/26335] [Medline: 35311684]
38. Wojcik S, Hughes A. Sizing Up Twitter Users. 2019. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2022-09-10]
39. Enli G, Simonsen C. 'Social media logic' meets professional norms: Twitter hashtags usage by journalists and politicians. Inf Commun Soc 2017 Mar 17;21(8):1081-1096. [doi: 10.1080/1369118x.2017.1301515]
40. Zappavigna M. Searchable Talk: Hashtags and Social Media Metadiscourse. London, UK: Bloomsbury; 2018.
41. Chun Tie Y, Birks M, Francis K. Grounded theory research: A design framework for novice researchers. SAGE Open Med 2019 Jan 02;7:2050312118822927 [FREE Full text] [doi: 10.1177/2050312118822927] [Medline: 30637106]
42. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. New York, NY: Routledge; 2017.
43. Neuendorf KA. The Content Analysis Guidebook. Thousand Oaks, CA: SAGE Publications; 2017.
44. Nicholson Foundation. Over Half of Tweets about ECigs Might Come from Bots. 2019. URL: https://thenicholsonfoundation.org/sites/default/files/PGP-TNFBot_Detection_Report.pdf [accessed 2022-09-10]
45. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
46. Varol O, Ferrara E, Davis C, Menczer F, Flammini A. Online human-bot interactions: detection, estimation, and characterization. 2017 Presented at: Eleventh International AAAI Conference on Web and Social Media; May 15-18, 2017; Montreal, Quebec, Canada.

47. Rauchfleisch A, Kaiser J. The false positive problem of automatic bot detection in social science research. PLoS One 2020 Oct 22;15(10):e0241045 [FREE Full text] [doi: 10.1371/journal.pone.0241045] [Medline: 33091067]
48. Majmundar A, Moran M. For or against tobacco control: sponsored tobacco advocacy messages on Facebook and Instagram. Nicotine Tob Res 2021 Oct 07;23(11):1861-1868. [doi: 10.1093/ntr/ntab111] [Medline: 34049405]
49. Bommele J, Hopman P, Walters BH, Geboers C, Croes E, Fong GT, et al. The double-edged relationship between COVID-19 stress and smoking: implications for smoking cessation. Tob Induc Dis 2020;18:63 [FREE Full text] [doi: 10.18332/tid/125580] [Medline: 32733178]
50. Klemperer EM, West JC, Peasley-Miklus C, Villanti AC. Change in tobacco and electronic cigarette use and motivation to quit in response to COVID-19. Nicotine Tob Res 2020 Aug 24;22(9):1662-1663 [FREE Full text] [doi: 10.1093/ntr/ntaa072] [Medline: 32343816]
51. Kreslake JM, Simard BJ, O'Connor KM, Patel M, Vallone DM, Hair EC. E-cigarette use among youths and young adults during the COVID-19 pandemic: United States, 2020. Am J Public Health 2021 Jun;111(6):1132-1140. [doi: 10.2105/ajph.2021.306210]
52. Kamiński M, Muth A, Bogdański P. Smoking, vaping, and tobacco industry during COVID-19 pandemic: Twitter data analysis. Cyberpsychol Behav Soc Netw 2020 Dec 30;23(12):811-817. [doi: 10.1089/cyber.2020.0384] [Medline: 32757951]
53. Perski O, Simons D, Shahab L, Brown J. Smoking, nicotine and COVID-19: triangulation of methods and pre-registration are required for robust causal inference. Nicotine Tob Res 2021 Oct 22:214. [doi: 10.1093/ntr/ntab214] [Medline: 34679177]
54. Maggi F, Rosellini A, Spezia PG, Focosi D, Macera L, Lai M, et al. Nicotine upregulates ACE2 expression and increases competence for SARS-CoV-2 in human pneumocytes. ERJ Open Res 2021 Apr;7(2):00713-2020 [FREE Full text] [doi: 10.1183/23120541.00713-2020] [Medline: 33850935]
55. Kim E, Ihm J. Online news sharing in the face of mixed audiences: context collapse, homophily, and types of social media. J Broadcast Electron Media 2020 Dec 09;64(5):756-776. [doi: 10.1080/08838151.2020.1835429]
56. Center for Countering Digital Hate. The Disinformation Dozen: Why Platforms Must Act on Twelve Leading Online Anti-Vaxxers. 2021. URL: https://counterhate.com/research/the-disinformation-dozen/ [accessed 2022-09-10]
57. Papacharissi Z. Affective Publics: Sentiment, Technology, and Politics. Oxford, UK: Oxford University Press; 2015.
58. Jeffares S. Interpreting Hashtag Politics: Policy Ideas in an Era of Social Media. Berlin: Springer; 2014.
59. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Mar 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]
60. Shao C, Hui P, Wang L, Jiang X, Flammini A, Menczer F, et al. Anatomy of an online misinformation network. PLoS One 2018 Apr 27;13(4):e0196087 [FREE Full text] [doi: 10.1371/journal.pone.0196087] [Medline: 29702657]

## Abbreviations

**API:** Application Programming Interface
**ICU:** intensive care unit
**RQ:** research question
**URL:** Uniform Resource Locator

*Edited by C Basch; submitted 15.06.22; peer-reviewed by JP Allem, M Nali; comments to author 21.07.22; revised version received 24.08.22; accepted 24.08.22; published 22.09.22.*

*Please cite as:*
*Silver N, Kierstead E, Kostygina G, Tran H, Briggs J, Emery S, Schillo B*
*The Influence of Provaping "Gatewatchers" on the Dissemination of COVID-19 Misinformation on Twitter: Analysis of Twitter Discourse Regarding Nicotine and the COVID-19 Pandemic*
*J Med Internet Res 2022;24(9):e40331*
*URL: https://www.jmir.org/2022/9/e40331*
*doi:10.2196/40331*
*PMID:36070451*

Original Paper

# Pain Reduction With an Immersive Digital Therapeutic Tool in Women Living With Endometriosis-Related Pelvic Pain: Randomized Controlled Trial

Benjamin Merlot[1], MD; Garance Dispersyn[2], PhD; Zoé Husson[1], MSc; Isabella Chanavaz-Lacheray[1], MD; Thomas Dennis[1], MD; Juliette Greco-Vuilloud[2], PhD; Maxime Fougère[2], PhD; Stéphane Potvin[3], PhD; Maryne Cotty-Eslous[2], MSc; Horace Roman[1], MD; Serge Marchand[2,4], PhD

[1]Franco European Multidisciplinary Endometriosis Institute (IFEMEndo), Bordeaux, France

[2]Lucine, Bordeaux, France

[3]Centre de Recherche de l'Institut Universitaire en Santé Mentale de Montréal, Montréal, QC, Canada

[4]Faculté de Médecine et des Sciences de la Santé, Centre de Recherche Clinique du Centre Hospitalier Universitaire de Sherbrooke, Université de Sherbrooke, Sherbrooke, QC, Canada

**Corresponding Author:**
Serge Marchand, PhD
Faculté de Médecine et des Sciences de la Santé
Centre de Recherche Clinique du Centre Hospitalier Universitaire de Sherbrooke
Université de Sherbrooke
3001, 12e Avenue Nord
Sherbrooke, QC, J1H5N4
Canada
Phone: 1 8193461110 ext 15889
Email: serge.marchand@usherbrooke.ca

## *Abstract*

**Background:** Chronic pelvic pain is a common and disabling condition in women living with endometriosis. Pharmacological and surgical treatments are not always effective at controlling pain and present important restrictions. Digital therapeutics (DTx) are emerging as major nonpharmacological alternatives that aim to extend the analgesic therapeutic arsenal of patients.

**Objective:** In this randomized controlled trial (RCT), we aimed to measure the immediate and 4-hour persisting effects of a single use 20-minute DTx (Endocare) on pain in women experiencing pelvic pain due to endometriosis.

**Methods:** A total of 45 women with endometriosis participated in a randomized controlled study comparing the analgesic effect of a single use of a virtual reality digital treatment named Endocare (n=23, 51%) to a 2D digital control (n=22, 49%). Perceived pain and pain relief were measured before the treatment and 15, 30, 45, 60, and 240 minutes after the end of the treatment.

**Results:** The clustered posttreatment pain was significantly reduced compared to the pretreatment for both Endocare and the control group (all $P<.01$). Endocare was significantly more effective than the control group (all $P<.01$). Endocare decreased the mean pain intensity from 6.0 (SD 1.31) before the treatment to 4.5 (SD 1.71) posttreatment, while the control only decreased it from 5.7 (SD 1.36) to 5.0 (SD 1.43). When comparing each posttreatment measures to the pretest, Endocare significantly reduced pain perception for all points in time up to 4 hours posttreatment. The differences did not reached significance for the control group. Moreover, Endocare was significantly superior to the control group 15, 30, and 45 minutes after the treatment (all $P<.001$). The mean perceived pain relief was significantly higher for Endocare at 28% (SD 2%) compared to the control, which was 15% (SD 1%) for all the posttreatment measurements (all $P>.05$).

**Conclusions:** Our study aimed to test the effects of a single use of a DTx treatment on reported pain at different time points in women diagnosed with endometriosis experiencing moderate-to-severe pelvic pain. Importantly, our results support that Endocare, a virtual reality immersive treatment, significantly reduce pain perception compared to a digital control in women living with endometriosis. Interestingly, we are the first to notice that the effect persisted up to 4 hours posttreatment.

**Trial Registration:** ClinicalTrials.gov NCT04650516; https://tinyurl.com/2a2eu9wv

***(J Med Internet Res 2022;24(9):e39531)*** doi:10.2196/39531

**KEYWORDS**

digital treatment; digital intervention; virtual reality; pelvic pain; endometriosis; digital therapeutics; chronic pain; randomized controlled trial; RCT; pain; women's health; eHealth; digital health; endometrium; pelvis; pelvic; efficacy; effectiveness; gynecology; gynecologist; sexual health; reproductive health

## Introduction

Endometriosis is characterized by lesions occurring outside the uterus whose appearance and behaviors are close to those of the endometrium mainly on the pelvic peritoneum, ovaries, and rectovaginal septum [1]. It might also reach deeper tissues, such as the rectum, bladder, ureters, colon, small bowel, diaphragm, or pelvic nerves. Endometriosis is a multifactorial disease resulting from the combined action of genetic and environmental factors [2,3] and is a major contributor to chronic pelvic pain (CPP) and infertility among women [4].

The mean prevalence of endometriosis in women with CPP has been estimated at 70% (SD 3%) [5], but it may vary from 2% to 93% depending on the study [5,6]. CPP is the main symptom of women with endometriosis [7]. The most common painful symptoms of endometriosis are dysmenorrhea and deep dyspareunia, appearing in nearly 80% and 30% of patients, respectively [1]. The pain, which can severely affect quality of life, can be constant or triggered by various conditions, such as menstruation [8].

Current management of endometriosis pain involves pharmacological (eg, hormonal therapy) and surgical treatments [9-11]. The goals of these medical therapies are multifold (eg, reduction of inflammation, inhibition of ovulation, suppression of menstruation) [1]. These are based on the concept that the response of the eutopic endometrium and endometriosis lesions is substantially similar.

Hormonal therapies or surgeries are not adapted for women willing to get pregnant. Therefore, the development of nonpharmacological alternatives for the management of endometriosis pain is critical to extend the treatment arsenal for women with pelvic-perineal pain and endometriosis [12]. Among them, pilot evidence asserts the benefits of mindfulness training (eg, meditation, breathing, music) on pain relief [13]. Various levels of quality of life (ie, physical, mental and social features) are affected in women living with endometriosis [14,15]. Therefore, additions to the treatment of psychological interventions (eg, hypnotherapy, cognitive behavioral therapy) have shown significant results for both physical and mental well-being in patients living with CPP and endometriosis [16,17].

In the past few decades, the emergence of digital therapeutics (DTx) aimed at the use of informatic tools to aid the diagnosis and therapy of various pathologies has led to the creation of many new therapeutic devices [18]. Recently, multiple meta-analyses have specifically supported the efficacy of virtual reality (VR) in several type of acute and chronic pain conditions [19-24].

On these bases, a new digital therapeutic approach, Endocare, was created by combining various therapeutic procedures based on several modalities in a VR environment, individually known to reduce pain [25-27]. Endocare therapeutic procedures comprise auditory (eg, alpha/theta binaural beats, nature-based sounds) and visual (eg, bilateral alternative stimulations) components associated with a 3D VR environment. To our knowledge, no studies have ever tested the effects of this type of nonpharmacological treatment on reported pain at different time points in patients diagnosed with endometriosis perceiving moderate-to-severe pelvic pain. In this randomized controlled trial (RCT), we hypothesized that a single use of the Endocare treatment would be able to significantly diminish the pain intensity in patients diagnosed with CPP associated with endometriosis.

## Methods

### Design and Setting

This was a randomized, controlled, comparative, open-label, 2-parallel-group interventional study comparing the effect of Endocare and a digital control on endometriosis-related pain after a single use. This RCT was conducted between December 2020 and May 2021 at the Franco-European Multidisciplinary Endometriosis Institute (IFEMEndo), Clinic Tivoli-Ducos in Bordeaux, France.

### Ethics Approval

This study was conducted in compliance with good clinical practice guidelines, the principles of the Declaration of Helsinki, and French laws and regulations. It was reviewed and approved on November 6, 2020, by the Comité de Protection des Personnes. All participants completed and signed the informed consent form before inclusion in the study and before any study-related procedure began. Before commencing participant enrollment, this study was registered on ClinicalTrials.gov (NCT04650516).

### Quality Assurance

Standard operating procedures were applied for the conduct and analysis of the clinical investigation. This study was monitored regularly according to the specifically designed monitoring plan. Furthermore, data were captured following the double entry procedure in the dedicated case report form and compared. Discrepancies were reviewed and corrected by a third entry clerk. A data validation document specifically written for this study aimed to list all data checks to be performed. A data manager programmed the checks with Ennov Clinsight, and the sponsor validated the checks. In case of inconsistencies, a query was edited during the quality analysis beyond data entry. The investigator or another authorized person from the clinical staff was asked to answer the query by confirming or correcting the data.

## Population

### Selection Criteria

The selection criteria were women over 18 years old with a magnetic resonance imagery (MRI) diagnosis of endometriosis who were willing to participate in the study and signed the informed consent form. All patients were recruited from a highly specialized center (ie, IFEMEndo). These patients experienced mostly severe pain related to their chronic pain condition and presented a long medical history concerning their endometriosis, which was often deep endometriosis.

### Inclusion Criteria

To be included in this study, the screened participants who met the selection criteria had to be living with moderate-to-severe endometriosis-related pain with a score ≥4 on an 11-point numerical rating scale (NRS) at the time of inclusion, a criterion shared among various studies [8].

### Exclusion Criteria

The exclusion criteria were women who (1) were pregnant or breastfeeding; (2) had consumed painkillers within 8 hours prior to inclusion; (3) were participating in an interventional study or had participated in an interventional study within 30 days before enrolment; (4) were employed by the investigator or study site, with direct involvement in the proposed study or other studies under the direction of that investigator or study site, as well as family members of the employees or the investigator; and (5) had a contraindication to Endocare or the digital control, such as severe visual, hearing or cognitive impairments, color blindness, photosensitivity, epilepsy, or motion sickness.

### Sample Size Determination

This study aimed to evaluate the short-term pain evolution after a single use of Endocare compared to a digital control in participants experiencing endometriosis-related pelvic pain. The primary end point was the mean pain intensity 60 minutes after the beginning of treatment (or last evaluation if the participant dropped out of the study or started any rescue medication). Gerlinger et al [28] determined that women who felt "minimally satisfied" in the management of their pain had a change of −19.5 mm (SD 14.3 mm) on a visual analogical scale. Hence, a 20-mm difference between Endocare and the digital control seemed like a reasonable clinical target and was taken into consideration for sample size calculation. We planned to measure a total of 40 participants, with 20 allocated to the intervention group and 20 to the control group, at 5 time points.

With a 2-sided 95% CI, the study achieved 81% power to detect a difference between the group means at the last time of 20 mm on a 100-mm VAS with a 25-mm standard deviation. The correlation between measurements within a participant was estimated at 0.500. A test based on a mixed-model analysis was anticipated at a significance level of 5%. Assuming that about 25% of women would not experience an endometriosis-related pain ≥4 on NRS at time of inclusion, we planned to screen 50 women. Consequently, at least 40 women were planned to be included in the study.

Finally, 46 participants were screened, of which 45 were included and randomized; 1 was a screen failure because she never went back for the study visit (Table 1). A total of 44 (97.78%) participants completed the entire study.

**Table 1.** Size of the sample.

| Sample | Value |
|---|---|
| Participants screened, n | 46 |
| Participants randomized, n (%) | 45[a] (100) |
| Participants who completed study, n (%) | 44[b] (97.8) |

[a] One participant was lost to follow-up between the screening visit and the randomization visit and their inclusion criteria could not be checked.

[b] The self-administered questionnaire of 1 participant was never received for analysis.

## Concomitant Medications

Pain medications were stopped at least 8 hours prior to participation. If pain persisted after treatment (Endocare or control), the patient could take their pain management rescue treatment, thus ending the data collection at that time.

## Treatment and Control

### Endocare

We developed the Endocare treatment (Lucine, Bordeaux, France) specifically for this study (Figure 1). It was displayed through a VR headset (Oculus Quest) with high-quality headphones (APK K-240-MKII). Endocare is a standalone medical software device comprised of an application stored in a VR headset that is intended to mitigate pain for people prone to endometriosis. Endocare offers a 20-minute treatment consisting of a combination of auditory (eg, alpha/theta binaural beats, nature-based sounds) and visual (eg, bilateral alternative stimulations consisting of a sphere appearing and moving on a horizontal axis) therapeutic procedures integrated in a 3D VR environment.

**Figure 1.** Visual display of the 3D virtual reality treatment (ie, Endocare) and the 2D tablet control.

### Control

The digital control was also developed by Lucine specifically for this study (Figure 1). The digital control program was displayed through a tablet (Samsung Galaxy Tab A) with high-quality headphones similar to those used in the Endocare group (APK K-240-MKII). The digital comparator was a 20-minute control with the same composition as the Endocare treatment (same context, environment, and duration) but without any immersive effects of the VR headset itself, nor the auditory (eg, alpha/theta binaural beats, nature-based sounds) and visual (eg, bilateral alternative stimulations) stimuli. A soundtrack composed of nature sounds related to the projected image.

## Procedures

Women living with endometriosis pain were recruited from IFEMEndo, a clinic that specializes in endometriosis. All patients were diagnosed with endometriosis by specialized gynecologists and medical procedures (eg, MRI). Recruitment was done when the patients came for their medical examination and were informed about this study. If the patients expressed interested in participating in this study, the research team offered them a new appointment at the clinic to receive the study treatment.

## On-site Visit

When patients returned to the clinic on the day of the inclusion visit, their pain level needed to be at least 4 on an 11-point numeric pain rating scale to participate in this study. Patients reporting a pain of <4 on this scale were considered as screen failures and withdrawn from the study. This pain intensity evaluation represented the pretreatment evaluation just before the use of Endocare or control and is considered the baseline (T0).

If the patient consented to participate and met the inclusion criteria, they were randomized to 1 of the 2 groups, Endocare or control. The treatment (either Endocare or control) consisted of a single use of the VR headset on the day of the visit, as presented in Table 2.

The research team accompanied the patient to distinct isolated rooms and then handed them either the headset (ie, Endocare) or the tablet (ie, control). Patients were then guided orally by the investigator for up to 5 minutes until the beginning of the treatment (ie, Endocare or control).

**Table 2.** Study flowchart.

| Study procedures | Screening | Care and follow-up | Follow-up until 240 minutes after treatment |
|---|---|---|---|
| Informed consent | ✓ | | |
| Eligibility criteria (without pain assessment) | ✓ | | |
| Demographics | ✓ | | |
| History and management of endometriosis | ✓ | | |
| Typology of pain crises | ✓ | | |
| Baseline pain assessment (T0) (just before the start of Endocare or control treatment) | | ✓ | |
| Assessment of general status | ✓ | ✓ | ✓ |
| Endocare or control treatment | | ✓ | |
| Pain evolution (pain relief and intensity) | | ✓[a] | ✓[b] |
| Satisfaction | | | ✓[b] |
| Concomitant treatments | | ✓[c] | ✓[c] |
| Adverse events | | ✓ | ✓ |

[a] Evaluation 15 minutes (T15), 30 minutes (T30), 45 minutes (T45), and 60 (T60) minutes posttreatment.

[b] Evaluation 240 minutes (T240) posttreatment.

[c] Includes rescue treatment if needed.

## Follow-up

Once the treatment was completed, each participant was asked to rate on a paper questionnaire their pain perception on an 11-point NRS (0: no pain, 10: unbearable pain) and their perceived pain relief on a 5-point categorical scale 15 minutes (T15), 30 minutes (T30), 45 minutes (T45), and 60 minutes (T60) after administration of the allocated study treatment. During the first hour, participants remained under the direct supervision of the study site staff. After the first hour, and if cleared by the study site staff, participants were free to leave the study site and go home to complete next assessment 240 minutes (T240) after the end of the treatment. Adverse events (AEs), if any, were collected during the entire duration of patients' participation. Before leaving the clinic, participants underwent a general health status check to ensure they were not experiencing any AEs and were able to go home.

At home, patients were asked to rate on a paper questionnaire their pain relief on a 5-point categorical scale and pain intensity based on the 11-point NRS at T240 or at the last time point if the participant dropped out the study or began taking a rescue medication before the end of the follow-up period.

## Measurements

### Pain Assessment and Pain Relief

Pain intensity was evaluated on an 11-point NRS at T0, T15, T30, T45, T60, and T240. This scale is the reference tool to assess pain in most clinical trials, and it has been used in various recent studies, including those related to endometriosis [29-31]. Notably, its use is recommended by the IMMPACT (Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials) guidelines [32].

Pain relief was evaluated on a 5-point categorical scale (0: no relief, 1: slight relief, 2: moderate relief, 3: lots of relief, and 4: complete relief) at T0, T15, T30, T45, T60, and T240.

If rescue medication was needed, the participant was asked to rate their pain intensity or pain relief just prior to the intake of the rescue medication. Study follow-up was terminated upon the intake of rescue medication.

### Patient Replacement

Patients who dropped out before the end of the 240-minute follow-up period were not replaced.

### Statistical Analyses

A linear mixed-model framework was used for statistical analysis (SPSS 2020, IBM Corp). The selected covariance matrix of the repeated measurements that better fitted the data based on the Akaike Information Criterion was the first-order ante-dependence. There were 3 explanatory fixed factors: group (Endocare, control), time (T0, T15, T30, T45, T60, T240), and group*time, as well as a random effect on the intercept of each participant. In the first analysis, time (fixed effect) was parametrized as baseline (T0) versus all other times (T15 to T240). The analysis consisted of comparing T0 with the 5 posttreatment data clustered for both groups and looking at the interaction term (group*time). A contrast analysis was also performed to verify the difference between baseline (T0) and posttreatment times (clustered) in the 2 separate groups. In the second analysis, time was parametrized as T0, T15, T30, T45, T60 and T240. A contrast analysis was also performed to test the difference between T0 and all other times (separated) in both groups. Bonferroni corrections were applied for multiple comparisons.

For pain relief, data were analyzed using R software. Because data did not satisfy to normality when assessed with the

Shapiro-Wilk normality test, Wilcoxon unilateral unpaired tests corrected for multiple comparisons with a false discovery rate (FDR) were used.

## Results

### Study Participants

Of the 45 women that participated in the study, 1 (2%) did not return the questionnaire. Among the patients included, 1 (2%) from the treatment group and 4 (9%) from the control group took rescue medication after T60 but before T240. Therefore, in accordance with the study design, no results at T240 were collected for these 5 patients. Participants from both groups were comparable in terms of age, height, and weight (Table 3), as well as in pain intensity at T0.

All participants were recruited from a clinic that specializes in endometriosis and were living with severe endometriosis-associated symptoms (Table 4). Of the participants, 90% (n=41) were living with CPP not related to menses, and the majority were living with dysmenorrhea, dysuria, and dyspareunia. Moreover, 77% (n=35) were living with deep infiltrating endometriosis, and 22% (n=10) had adenomyosis.

Most of the patients were taking different classes of medications for their endometriosis condition, including hormones, analgesics, and antidepressants (Table 4). Since this study was based on an add-on protocol, participants were authorized to continue all their medications, except for pain drugs before the beginning of the testing. The participants kept these analgesics as rescue medication after the Endocare or control treatment if needed.

**Table 3.** Participant demographics.

| Characteristics | Total (N=45) | Endocare (n=23) | Control (n=22) |
|---|---|---|---|
| **Age (years)** | | | |
| M[a] (missing[b]) | 45 (0) | 23 (0) | 22 (0) |
| Mean (SD) | 32.7 (8.02) | 32.2 (8.02) | 33.2 (8.12) |
| Median | 31 | 32 | 30.5 |
| Min[c] | 21 | 21 | 21 |
| Max[d] | 53 | 53 | 51 |
| **Height (cm)** | | | |
| M (missing) | 44 (1) | 22 (1) | 22 (0) |
| Mean (SD) | 162.7 (7.22) | 162.6 (6.72) | 162.8 (7.84) |
| Median | 165 | 165 | 164 |
| Min | 148 | 148 | 150 |
| Max | 175 | 175 | 175 |
| **Weight (kg)** | | | |
| M (missing) | 44 (1) | 22 (1) | 22 (0) |
| Mean (SD) | 63.8 (12.51) | 67.2 (12.46) | 60.4 (11.88) |
| Median | 60 | 63.5 | 57 |
| Min | 45 | 50 | 45 |
| Max | 95 | 95 | 92 |

[a] M: total number of cases.

[b] Missing: total number of cases with missing data.

[c] Min: minimum.

[d] Max: maximum.

**Table 4.** Participants' history of endometriosis.

| Variable | Total (N=45) | Endocare (n=23) | Control (n=22) |
|---|---|---|---|
| **Time since endometriosis diagnosis (months)** | | | |
| M[a] (missing[b]) | 45 (0) | 23 (0) | 22 (0) |
| Mean (SD) | 45.2 (54.9) | 43.4 (59.3) | 47.2 (51.1) |
| Median | 23.8 | 16.6 | 23.8 |
| Min[c] | 2.2 | 2.5 | 2.2 |
| Max[d] | 277.8 | 277.8 | 199.7 |
| **Type of endometriosis, n (%)** | | | |
| Superficial peritonea | 10 (22.2) | 6 (26.1) | 4 (18.2) |
| Deep infiltrating endometriosis | 34 (75.6) | 17 (73.9) | 17 (77.3) |
| Digestive locations | 1 (2.2) | 1 (4.3) | 0 (0) |
| Other | 2 (4.4) | 1 (4.3) | 1(4.5) |
| Presence of adenomyosis, n (%) | 11 (24.4) | 6 (26.1) | 5 (22.7) |
| History of surgical management of endometriosis, n (%) | 31 (68.9) | 14 (60.9) | 17 (77.3) |
| **Time from endometriosis diagnosis to surgery (months)** | | | |
| M (missing) | 31 (14) | 14 (9) | 17 (5) |
| Mean (SD) | 40.6 (43.9) | 40.5 (36.5) | 40.6 (50.2) |
| Median | 24.4 | 38.8 | 23.8 |
| Min | 0.1 | 2.1 | 0.1 |
| Max | 199.7 | 136.0 | 199.7 |
| **Symptom resolution following surgery, n (%)** | | | |
| Yes | 13 (41.9) | 3 (21.4) | 10 (58.8) |
| No | 18 (58.1) | 11 (78.6) | 7 (41.2) |
| **Time from surgery to symptom recurrence (months)** | | | |
| M (missing) | 13 (0) | 3 (0) | 10 (0) |
| Mean (SD) | 35.1 (55.2) | 9.5 (15.5) | 42.8 (61.1) |
| Median | 21.1 | 1.7 | 22.0 |
| **Current management of endometriosis, n (%)** | | | |
| **Hormone-based therapy** | | | |
| None | 10 (22.2) | 6 (26.1) | 4 (18.2) |
| Combined hormonal contraceptives | 14 (31.1) | 7 (30.4) | 7 (31.8) |
| Progestogens | 18 (40.0) | 9 (39.1) | 9 (40.9) |
| Gonadotrophin-releasing hormone agonists | 1 (2.2) | 0 (0) | 1 (4.5) |
| Other | 2 (4.4) | 1 (4.3) | 1 (4.5) |
| Chronic analgesic treatment | 26 (57.8) | 14 (60.9) | 12 (54.5) |
| Other treatment | 16 (35.6) | 7 (30.4) | 9 (40.9) |
| **Usual pain symptoms, n (%)** | | | |
| Chronic pelvic pain not related to menses | 41 (91.1) | 21 (91.3) | 20 (90.9) |
| Dysmenorrhea | 24 (53.3) | 12 (52.2) | 12 (54.5) |
| Dysuria | 14 (31.1) | 6 (26.1) | 8 (36.4) |
| Dyschesia | 1 (2.2) | 1 (4.3) | 0 (0) |
| Dyspaneuria | 29 (64.4) | 15 (65.2) | 14 (63.6) |

[a] M: total number of cases.

[b] Missing: total number of cases with missing data.

[c] Min: minimum.

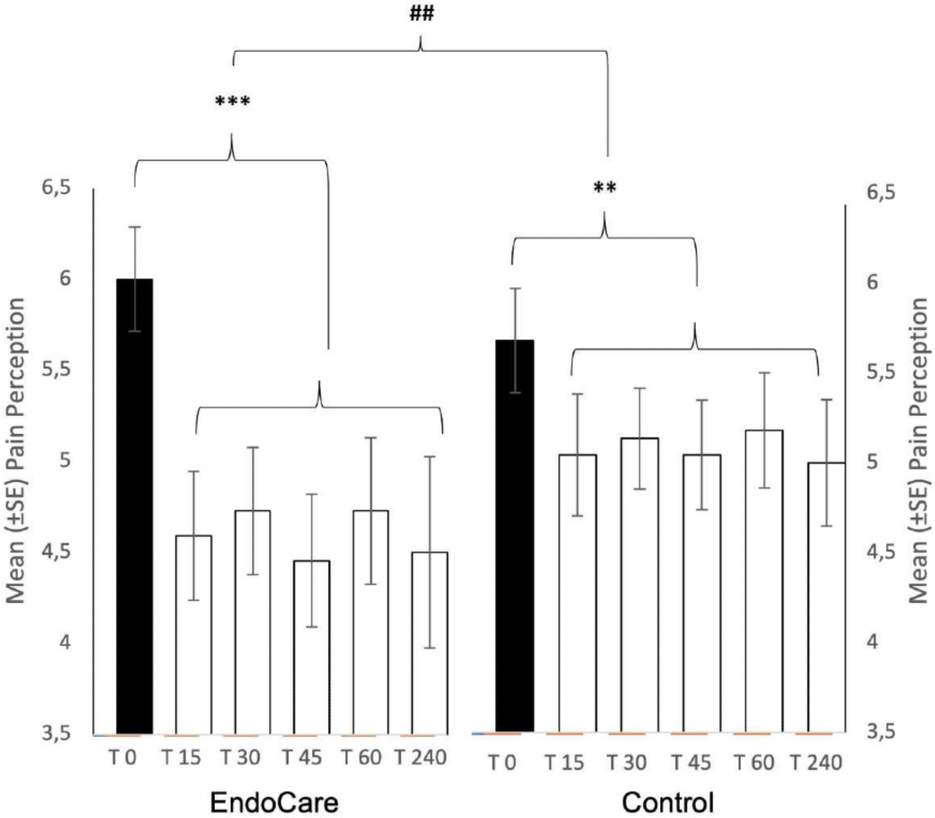
[d] Max: maximum.

## Pain Perception

### Difference in Pain Perception Between Baseline and the 5 Clustered Posttreatment Measurements

We first analyzed the differences between the baseline pain (T0) and the 5 clustered posttreatment measurements covering 15 minutes to 4 hours after the treatment (ie, T15, T30, T45, T60, and T240) in both groups using a linear mixed model (Figure 2). We did not observe any effect of the group (Endocare, control; $F_{40,712}$=.060; $P$=.807), but we did observe an effect of the time (T0; $F_{43,094}$=44.179; $P$<.001). Moreover, we found an interaction for group*time ($F_{43,094}$=7.343; $P$=.010), indicating that the mean reduction of pain through time (T0) was significantly different between each group (ie, Endocare, control). Contrast analysis revealed that the mean reduction of pain was greater in the Endocare group (1.58; $t_{42,926}$=6.624; $P$<.001) than in the control group (0.38; $t_{43,252}$=2.781; $P$=.008).

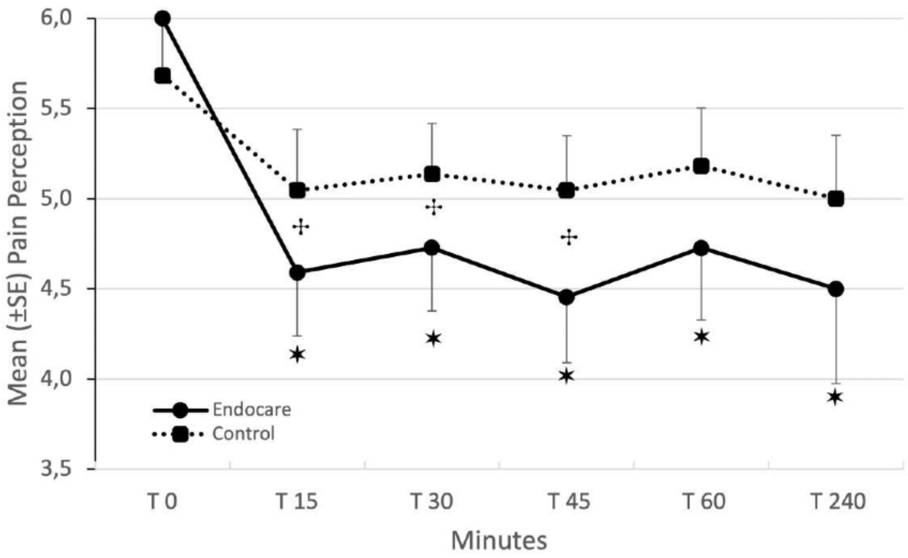
**Figure 2.** Difference in pain perception between baseline and the 5 clustered posttreatment measurements. Mean pain perception (± standard error) in both groups (ie, Endocare, control) at baseline (black: T0) and the 5 clustered posttreatment measurements (white: T15, T30, T45, T60, and T240). ##$P$<.01 (interaction group*time); **$P$<.01 (mixed model) ***$P$<.001 (mixed model).



### Difference in Pain Perception Between Baseline and Each of the 5 Posttreatment Measurements

We then aimed to analyze the differences between the baseline (T0) and each of the 5 posttreatment measurements separately (ie, T15, T30, T45, T60, or T240) in both groups using a mixed model (Figure 3). We did not observe any effect of the group ($F_{41,122}$=0.716; $P$=.402), but we did observe an effect of the time ($F_{57,235}$=10.066; $P$<.001). Moreover, we did not find an interaction of group*time ($F_{57,235}$=1.618; $P$=.170), likely due to a weak statistical power. Contrast analysis revealed group*time interactions at T15 ($t_{42,000}$=2.211; $P$=.033), T30 ($t_{46,241}$=2.226; $P$=.031), and T45 ($t_{41,389}$=2.53; $P$=.015), a tendency for T60 ($t_{38,964}$=1.946; $P$=.059), and none for T240 ($t_{37,872}$=1.618; $P$=.114). This indicates that the mean reduction of pain through time was significantly different between each group (ie, Endocare, control) from T15 to T45, but it was not significantly different from T60 to T240. For the Endocare group, the pain perception reduction was significant in each of the 5 posttreatment measurements (Figure 2, all $P$<.05). For the control group, except for a tendency at T15 ($t_{42,000}$=2.576; $P$=.068), no significance was reached relative to the pain perception reduction in each of the 5 posttreatment measurements (all $P$>.05).

**Figure 3.** Difference in pain perception before and each of the 5 posttreatments measurements. Mean pain perception (± standard error) in both groups (bold line: Endocare; dash-line: control) at baseline (T0) and each of the 5 posttreatment measurements (T15, T30, T45, T60, or T240). +*P*<.05 (*t* test); **P*<.05 (mixed model).



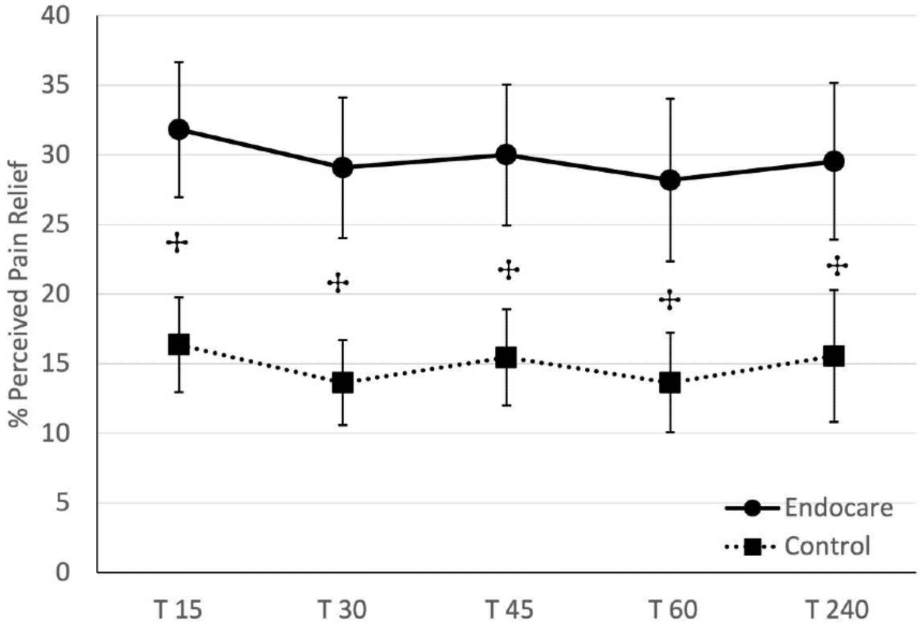
### *Covariance For Adenomyosis and Surgery*

A total of 31 (70%) participants had surgery related to endometriosis (14 in the Endocare group and 17 in the control), and 11 (25%) suffered from adenomyosis (6 in the Endocare group and 5 in the control). Thus, we introduced 2 covariates, surgery and adenomyosis, into the linear mixed model. Neither surgery ($F_{41,800}$=.009; *P*=.924) nor adenomyosis ($F_{41,782}$=.001; *P*=.982) changed the results. This indicates that the effect of the treatment on pain was independent of these 2 conditions.

### *Difference in Pain Relief at Each of the 5 Posttreatment Measurements*

At each posttreatment measurements, participants were asked to measure their perceived pain relief on a 5-point numerical scale from no relief to total relief (Figure 4). The mean pain relief reported was 28% (SD 24.28%) for Endocare and 15% (SD 16.34%) for the control group. Nonparametric Wilcoxon unilateral unpaired tests corrected for multiple comparisons with FDR were performed. All posttreatment measurements present a significantly higher pain relief score for Endocare compared to the control group (all *P*<.05).

**Figure 4.** Difference in pain relief at each of the 5 posttreatment measurements. Percentage of pain relief (± standard error of the mean) in both groups (circles and bold line: Endocare; squares and dash-line: control) at each of the 5 posttreatment measurements (T15, T30, T45, T60, or T240). +*P*<.05 (Wilcoxon).



### *Maximum Reduction in Pain Assessed by NRS*

Finally, we analyzed the maximum reduction of pain intensity in both groups. The mean maximum effect was 42% (95% CI 30.82-53.18) for Endocare and 22% (95% CI 15.38-28.53) for the control group. The maximum effect was significantly higher for the treatment group (Endocare; Cochran *t*-test, *P*=.004).

### *Adverse Events*

Seven (15%) participants reported mild-to-moderate AEs, of which 4 (8%) were evaluated as probably unrelated and 3 (6%) as possibly related to the Endocare treatment. The 3 possibly related events were within the treatment group and described as a mild headache and nausea related to motion sickness.

# Discussion

## Principal Results

In this RCT, we aimed to measure the immediate and 4 hours persisting effects of a single use of a 20-minute DTx (Endocare) on pain in women living with pelvic pain related to endometriosis. We showed that Endocare was able to significantly reduce the overall pain perception when comparing the pain at baseline (T0) with the combined 5 posttreatment measurements from 15 minutes (T15) to 4 hours (T240).

We found that our digital control was also able to significantly reduce the overall pain perception but significantly less than the Endocare treatment. Next, we wanted to evaluate whether the pain reduction was significant at each posttreatment measurement (ie, T15, T30, T45, T60, and T240) compared to baseline (T0) in each group. We showed that this was the case for the Endocare group but not for the control group, thus confirming the analgesic effect of the Endocare treatment. To our knowledge, our study is the first to demonstrate an effect on pain reduction in women living with moderate-to-severe pelvic pain due to endometriosis by using a nonpharmacological treatment that combined VR with visual and auditory stimuli.

## Comparison With Prior Work

Numerous studies have shown that VR is an effective way to reduce acute pain intensity, especially pain experienced during medical procedures, by burn victims, or by women during childbirth [24,33-36]. The vast majority of the studies on acute pain only measured pain intensity directly after the treatment, within minutes, or up to 1 hour [34,35,37,38]. Some studies also looked at the effects of VR on chronic pain [24,36]. For instance, studies on patients living with neuropathic, musculoskeletal, phantom limb, or indefinite chronic pain showed reduced pain intensity during or immediately after VR application [39-43]. More recently, a study also reported that 56 days of VR was effective at reducing chronic low back pain throughout the entire duration of the treatment [8]. However, there was no precise indication of how long participants rated their pain after the treatment intake during each of the 56 days. In our study, we measured pain intensity in a short-term period, showing that VR can reduce pain intensity for up to 4 hours, which is comparable to some analgesics. Indeed, a systematic review noted multiple analgesics for their ability to relieve acute postoperative pain by half over a 4-6-hour period [44]. In our study, to potentialize VR effects, we chose various auditory (eg, alpha/theta binaural beats, nature-based sounds) and visual (eg, bilateral alternative stimulations) therapeutic procedures known to relieve pain in different conditions (eg, premenstrual pain, medical imaging procedures, experimental pain, or caesarean section [25-27]). Lucine aimed to combine these stimuli into the DTx treatment Endocare to act on different components of pain. Among the auditory stimuli used in the treatment, binaural beats are associated with a calm and positive affect and are known to reduce stress and anxiety [45,46], contributing to reduced pain perception. Studies also reported that nature-based sounds can promote relief, concentration, and asleep [47], especially by masking environmental noise [48,49] or by amplifying slow waves sleep [50,51]. Therefore, these auditory stimuli can help people with CPP and endometriosis relax and feel reduced pain. Listening to nature sounds is also used in the medical context to reduce pain during surgeries, such as cesarean section, or for intensive care patients [27,52]. Moreover, a systematic review indicated that visual bilateral alternative stimulations can effectively reduce chronic pain [53].

Thus, the combination of these stimulations potentialized with the VR experience could lead to the decrease in pain observed in the Endocare group immediately after and up to 4 hours posttreatment.

In the field of VR research, while some studies do not use a control group [40,42], others compare the effects of VR with medication intake [54] or with totally different controls, such as closing the eyes or a distraction selected by the control group itself (eg, reading, meditating) [35,39,41]. This can be problematic since it is difficult to conclude on the use of VR or a simple headset effect. To our knowledge, only 1 study used a sham VR group, with noninteractive 2D nature scenes, and reported positive outcomes in both groups with reductions in pain and all domains of pain-related interference [8]. In this study, the sham group used the same VR apparatus as the other group. However, the environment of the 2 groups was not identical: the treatment group had a visual display skill-based interactive 3D environment, while the sham group had a noninteractive 2D nature scene. We purport that this type of control is closer to the treatment and thus easier to be double-blinded. Nevertheless, we could not definitively conclude on whether our results were due to the stimulations or the headset itself, since the immersive effect of the headset could be considered as part of the treatment [55].

In this study, we chose to use a digital control program displayed through a 2D tablet with a headset, allowing us to control for the Endocare treatment and not solely the VR apparatus or the stimuli. Indeed, our digital control contained the same composition as the Endocare treatment (ie, context, environment, duration) without the immersive aspect of VR or the auditory and visual stimuli of the treatment. However, to preserve an immersive-like session, we chose to keep a soundtrack composed of nature sounds related to the projected image in the control group, which could potentially explain the positive outcome observed. This slight decrease in pain intensity observed in the control group indicates that our control is relevant.

## Limitations

One limitation of our study may be the population, as we decided to work with patients living with mostly severe pain related to a complex and chronic pain condition, which could represent a selection bias. We likely could have had slightly better treatment results with less severe pain conditions. However, to assess this potential selection bias and analyze whether the beneficial effect would be lessened in populations with complex pathologies related to endometriosis, we covariated for surgery and adenomyosis, 2 conditions with potentially more pain and less responses to the treatment. We found that the results were equivalent, suggesting a similar effect in all subpopulations, including those with a more complex pathophysiology. Nevertheless, an effect in the range of 30% for Endocare can ultimately be considered moderately significant according to the IMMPACT guidelines [56]. This can be explained by the design of our study itself. Indeed, our relatively small sample size, or perhaps the unique use of the treatment, could have lessened the effect. Nevertheless, it is important to highlight the fact that we reached significance even with this design. Future studies should aim to confirm these results with a larger population size and repeated use of the treatment.

## Conclusions

In conclusion, this pilot study on the Endocare treatment shows encouraging results for developing a digital therapy to relieve patients of pelvic and perineal pain associated with endometriosis. Moreover, Endocare treatment can be a great alternative to hormonal treatment or surgery for women who wish to get pregnant. These results will be further investigated in a second study evaluating the analgesic effects of the repeated use of Endocare at home with a larger population of women living with chronic pelvic pain associated with endometriosis. In the future, Endocare could benefit patients diagnosed with endometriosis during their everyday life to reduce the acute and chronic pain encountered in this pathology, thus improving their quality of life.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 381 KB - jmir_v24i9e39531_app1.pdf ]

## References

1. Vercellini P, Viganò P, Somigliana E, Fedele L. Endometriosis: pathogenesis and treatment. Nat Rev Endocrinol 2014 May;10(5):261-275. [doi: 10.1038/nrendo.2013.255] [Medline: 24366116]
2. Saunders PTK, Horne AW. Endometriosis: Etiology, pathobiology, and therapeutic prospects. Cell 2021 May 27;184(11):2807-2824 [FREE Full text] [doi: 10.1016/j.cell.2021.04.041] [Medline: 34048704]
3. Taylor HS, Kotlyar AM, Flores VA. Endometriosis is a chronic systemic disease: clinical challenges and novel innovations. Lancet 2021 Feb 27;397(10276):839-852. [doi: 10.1016/S0140-6736(21)00389-5] [Medline: 33640070]
4. Giudice LC. Clinical practice. Endometriosis. N Engl J Med 2010 Jun 24;362(25):2389-2398 [FREE Full text] [doi: 10.1056/NEJMcp1000274] [Medline: 20573927]
5. Tirlapur SA, Kuhrt K, Chaliha C, Ball E, Meads C, Khan KS. The 'evil twin syndrome' in chronic pelvic pain: a systematic review of prevalence studies of bladder pain syndrome and endometriosis. Int J Surg 2013 May;11(3):233-237 [FREE Full text] [doi: 10.1016/j.ijsu.2013.02.003] [Medline: 23419614]
6. Guo S, Wang Y. The prevalence of endometriosis in women with chronic pelvic pain. Gynecol Obstet Invest 2006;62(3):121-130. [doi: 10.1159/000093019] [Medline: 16675908]
7. Dydyk A, Gupta N. Chronic Pelvic Pain. Treasure Islands, FL: StatPearls Publishing; May 29, 2022.
8. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. J Med Internet Res 2021 Feb 22;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]
9. Carbone MG, Campo G, Papaleo E, Marazziti D, Maremmani I. The Importance of a Multi-Disciplinary Approach to the Endometriotic Patients: The Relationship between Endometriosis and Psychic Vulnerability. J Clin Med 2021 Apr 10;10(8) [FREE Full text] [doi: 10.3390/jcm10081616] [Medline: 33920306]
10. Mechsner S. Management of endometriosis pain : Stage-based treatment strategies and clinical experience. Schmerz 2021 Jun;35(3):159-171. [doi: 10.1007/s00482-021-00543-8] [Medline: 33704582]
11. Vercellini P, Buggio L, Frattaruolo MP, Borghi A, Dridi D, Somigliana E. Medical treatment of endometriosis-related pain. Best Pract Res Clin Obstet Gynaecol 2018 Aug;51:68-91. [doi: 10.1016/j.bpobgyn.2018.01.015] [Medline: 29530425]
12. Wattier J. Conventional analgesics and non-pharmacological multidisciplinary therapeutic treatment in endometriosis: CNGOF-HAS Endometriosis Guidelines. Gynecol Obstet Fertil Senol 2018 Mar;46(3):248-255. [doi: 10.1016/j.gofs.2018.02.002] [Medline: 29510963]

13. Kold M, Hansen T, Vedsted-Hansen H, Forman A. Mindfulness-based psychological intervention for coping with pain in endometriosis. Nord Psychol 2012 Aug 31;64(1):2-16. [doi: 10.1080/19012276.2012.693727]
14. Jia S, Leng J, Shi J, Sun P, Lang J. Health-related quality of life in women with endometriosis: a systematic review. J Ovarian Res 2012 Oct 18;5(1):29 [FREE Full text] [doi: 10.1186/1757-2215-5-29] [Medline: 23078813]
15. Kennedy S, Bergqvist A, Chapron C, D'Hooghe T, Dunselman G, Greb R, et al. ESHRE guideline for the diagnosis and treatment of endometriosis. Hum Reprod 2005 Oct;20(10):2698-2704. [doi: 10.1093/humrep/dei135] [Medline: 15980014]
16. Ball E, Khan KS. Recent advances in understanding and managing chronic pelvic pain in women with special consideration to endometriosis. F1000Res 2020 Feb 4;9:83 [FREE Full text] [doi: 10.12688/f1000research.20750.1] [Medline: 32089831]
17. Meissner K, Schweizer-Arau A, Limmer A, Preibisch C, Popovici R, Lange I, et al. Psychotherapy with somatosensory stimulation for endometriosis-associated pain: a randomized controlled trial. Obstet Gynecol 2016 Nov;128(5):1134-1142. [doi: 10.1097/AOG.0000000000001691] [Medline: 27741200]
18. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. J Family Med Prim Care 2020 May;9(5):2207-2213 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_105_20] [Medline: 32754475]
19. Bordeleau M, Stamenkovic A, Tardif P, Thomas J. The Use of Virtual Reality in Back Pain Rehabilitation: A Systematic Review and Meta-Analysis. J Pain 2022 Feb;23(2):175-195. [doi: 10.1016/j.jpain.2021.08.001] [Medline: 34425250]
20. Chow H, Hon J, Chua W, Chuan A. Effect of virtual reality therapy in reducing pain and anxiety for cancer-related medical procedures: a systematic narrative review. J Pain Symptom Manage 2021 Feb;61(2):384-394. [doi: 10.1016/j.jpainsymman.2020.08.016] [Medline: 32822755]
21. Chuan A, Zhou JJ, Hou RM, Stevens CJ, Bogdanovych A. Virtual reality for acute and chronic pain management in adult patients: a narrative review. Anaesthesia 2021 May 27;76(5):695-704 [FREE Full text] [doi: 10.1111/anae.15202] [Medline: 32720308]
22. Ding L, Hua H, Zhu H, Zhu S, Lu J, Zhao K, et al. Effects of virtual reality on relieving postoperative pain in surgical patients: A systematic review and meta-analysis. Int J Surg 2020 Oct;82:87-94. [doi: 10.1016/j.ijsu.2020.08.033] [Medline: 32882400]
23. López-Valverde N, Muriel-Fernández J, López-Valverde A, Valero-Juan LF, Ramírez JM, Flores-Fraile J, et al. Use of virtual reality for the management of anxiety and pain in dental treatments: systematic review and meta-analysis. J Clin Med 2020 Sep 24;9(10):3086 [FREE Full text] [doi: 10.3390/jcm9103086] [Medline: 32987885]
24. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. JPR 2019 Jul;Volume 12:2053-2085. [doi: 10.2147/jpr.s200498]
25. Boyle Y, Bentley D, Watson A, Jones A. Acoustic noise in functional magnetic resonance imaging reduces pain unpleasantness ratings. Neuroimage 2006 Jul 01;31(3):1278-1283. [doi: 10.1016/j.neuroimage.2006.01.025] [Medline: 16517183]
26. Ecsy K, Jones A, Brown C. Alpha-range visual and auditory stimulation reduces the perception of pain. Eur J Pain 2017 Mar 02;21(3):562-572. [doi: 10.1002/ejp.960] [Medline: 27807916]
27. Farzaneh M, Abbasijahromi A, Saadatmand V, Parandavar N, Dowlatkhah HR, Bahmanjahromi A. Comparative effect of nature-based sounds intervention and headphones intervention on pain severity after cesarean section: a prospective double-blind randomized trial. Anesth Pain Med 2019 Apr;9(2):e67835 [FREE Full text] [doi: 10.5812/aapm.67835] [Medline: 31341820]
28. Gerlinger C, Schumacher U, Faustmann T, Colligs A, Schmitz H, Seitz C. Defining a minimal clinically important difference for endometriosis-associated pelvic pain measured on a visual analog scale: analyses of two placebo-controlled, randomized trials. Health Qual Life Outcomes 2010 Nov 24;8:138 [FREE Full text] [doi: 10.1186/1477-7525-8-138] [Medline: 21106059]
29. Facchin F, Barbara G, Saita E, Mosconi P, Roberto A, Fedele L, et al. Impact of endometriosis on quality of life and mental health: pelvic pain makes the difference. J Psychosom Obstet Gynaecol 2015;36(4):135-141. [doi: 10.3109/0167482X.2015.1074173] [Medline: 26328618]
30. Arendsen LJ, Henshaw J, Brown CA, Sivan M, Taylor JR, Trujillo-Barreto NJ, et al. Entraining alpha activity using visual stimulation in patients with chronic musculoskeletal pain: a feasibility study. Front Neurosci 2020 Aug 20;14:828 [FREE Full text] [doi: 10.3389/fnins.2020.00828] [Medline: 32973429]
31. Terho A, Puhto T, Laru J, Uimari O, Ohtonen P, Rautio T, et al. Laparoscopically guided transversus abdominis plane block versus local wound analgesia in laparoscopic surgery for peritoneal endometriosis: study protocol for a prospective randomized controlled double-blinded LTAP-trial. Trials 2022 Jan 18;23(1):55 [FREE Full text] [doi: 10.1186/s13063-022-06004-6] [Medline: 35042563]
32. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
33. Bermo M, Patterson D, Sharar S, Hoffman H, Lewis D. Virtual reality to relieve pain in burn patients undergoing imaging and treatment. Top Magn Reson Imaging 2020 Aug;29(4):203-208. [doi: 10.1097/RMR.0000000000000248] [Medline: 32511197]
34. Frey DP, Bauer ME, Bell CL, Low LK, Hassett AL, Cassidy RB, et al. Virtual reality analgesia in labor. Anesth Analg 2019;128(6):e93-e96. [doi: 10.1213/ane.0000000000003649]

35. Guo C, Deng H, Yang J. Effect of virtual reality distraction on pain among patients with hand injury undergoing dressing change. J Clin Nurs 2015 Jan 04;24(1-2):115-120. [doi: 10.1111/jocn.12626] [Medline: 24899241]
36. Pourmand A, Davis S, Marchak A, Whiteside T, Sikka N. Virtual reality as a clinical tool for pain management. Curr Pain Headache Rep 2018 Jun 15;22(8):53. [doi: 10.1007/s11916-018-0708-2] [Medline: 29904806]
37. Mosso-Vázquez JL, Gao K, Wiederhold BK, Wiederhold MD. Virtual reality for pain management in cardiac surgery. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):371-378 [FREE Full text] [doi: 10.1089/cyber.2014.0198] [Medline: 24892200]
38. JahaniShoorab N, Ebrahimzadeh Zagami S, Nahvi A, Mazluom SR, Golmakani N, Talebii M, et al. The effect of virtual reality on pain in primiparity women during episiotomy repair: a randomize clinical trial. Iran J Med Sci 2015 May;40(3):219-224 [FREE Full text] [Medline: 25999621]
39. Jin W, Choo A, Gromala D, Shaw C, Squire P. A virtual reality game for chronic pain management: a randomized, controlled clinical study. Stud Health Technol Inform 2016;220:154-160. [Medline: 27046570]
40. Jones T, Moore T, Choo J. The impact of virtual reality on chronic pain. PLoS One 2016 Dec 20;11(12):e0167523 [FREE Full text] [doi: 10.1371/journal.pone.0167523] [Medline: 27997539]
41. Wiederhold BK, Gao K, Sulea C, Wiederhold MD. Virtual reality as a distraction technique in chronic pain patients. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):346-352 [FREE Full text] [doi: 10.1089/cyber.2014.0207] [Medline: 24892196]
42. Osumi M, Ichinose A, Sumitani M, Wake N, Sano Y, Yozu A, et al. Restoring movement representation and alleviating phantom limb pain through short-term neurorehabilitation with a virtual reality system. Eur J Pain 2017 Jan 05;21(1):140-147. [doi: 10.1002/ejp.910] [Medline: 27378656]
43. Ichinose A, Sano Y, Osumi M, Sumitani M, Kumagaya S, Kuniyoshi Y. Somatosensory feedback to the cheek during virtual visual feedback therapy enhances pain alleviation for phantom arms. Neurorehabil Neural Repair 2017 Aug 08;31(8):717-725. [doi: 10.1177/1545968317718268] [Medline: 28691602]
44. Moore R, Derry S, McQuay HJ, Wiffen P. Single dose oral analgesics for acute postoperative pain in adults. Cochrane Database Syst Rev 2011 Sep 07(9):CD008659 [FREE Full text] [doi: 10.1002/14651858.CD008659.pub2] [Medline: 21901726]
45. Casciaro F, Laterza V, Conte S, Pieralice M, Federici A, Todarello O, et al. Alpha-rhythm stimulation using brain entrainment enhances heart rate variability in subjects with reduced HRV. WJNS 2013;03(04):213-220. [doi: 10.4236/wjns.2013.34028]
46. Le Scouarnec RP, Poirier RM, Owens JE, Gauthier J, Taylor AG, Foresman PA. Use of binaural beat tapes for treatment of anxiety: a pilot study of tape preference and outcomes. Altern Ther Health Med 2001 Jan;7(1):58-63. [Medline: 11191043]
47. Pickens TA, Khan SP, Berlau DJ. White noise as a possible therapeutic option for children with ADHD. Complement Ther Med 2019 Feb;42:151-155. [doi: 10.1016/j.ctim.2018.11.012] [Medline: 30670235]
48. Farokhnezhad Afshar P, Bahramnezhad F, Asgari P, Shiri M. Effect of white noise on sleep in patients admitted to a coronary care. J Caring Sci 2016;5(2):103-109 [FREE Full text] [doi: 10.15171/jcs.2016.011] [Medline: 27354974]
49. Stanchina ML, Abu-Hijleh M, Chaudhry BK, Carlisle CC, Millman RP. The influence of white noise on sleep in subjects exposed to ICU noise. Sleep Med 2005 Sep;6(5):423-428. [doi: 10.1016/j.sleep.2004.12.004] [Medline: 16139772]
50. Papalambros NA, Santostasi G, Malkani RG, Braun R, Weintraub S, Paller KA, et al. Acoustic enhancement of sleep slow oscillations and concomitant memory improvement in older adults. Front Hum Neurosci 2017 Mar 08;11:109 [FREE Full text] [doi: 10.3389/fnhum.2017.00109] [Medline: 28337134]
51. Zhou J, Liu D, Li X, Ma J, Zhang J, Fang J. Pink noise: effect on complexity synchronization of brain activity and sleep consolidation. J Theor Biol 2012 Aug 07;306:68-72. [doi: 10.1016/j.jtbi.2012.04.006] [Medline: 22726808]
52. Saadatmand V, Rejeh N, Heravi-Karimooi M, Tadrisi SD, Zayeri F, Vaismoradi M, et al. Effect of nature-based sounds' intervention on agitation, anxiety, and stress in patients under mechanical ventilator support: a randomised controlled trial. Int J Nurs Stud 2013 Jul;50(7):895-904. [doi: 10.1016/j.ijnurstu.2012.11.018] [Medline: 23245705]
53. Tesarz J, Leisner S, Gerhardt A, Janke S, Seidler GH, Eich W, et al. Effects of eye movement desensitization and reprocessing (EMDR) treatment in chronic pain patients: a systematic review. Pain Med 2014 Feb 05;15(2):247-263. [doi: 10.1111/pme.12303] [Medline: 24308821]
54. Bani Mohammad E, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: A randomized control trial. In: Palliative & Supportive Care. Cambridge, UK: Cambridge University Press; Sep 10, 2018:A.
55. Persky S, Lewis M. Advancing science and practice using immersive virtual reality: what behavioral medicine has to offer. Transl Behav Med 2019 Nov 25;9(6):1040-1046 [FREE Full text] [doi: 10.1093/tbm/ibz068] [Medline: 31116851]
56. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008 Feb;9(2):105-121. [doi: 10.1016/j.jpain.2007.09.005] [Medline: 18055266]

---

## Abbreviations

**AE:** adverse event

**CPP:** chronic pelvic pain
**DTx:** digital therapeutics
**FDR:** false discovery rate
**IFEMEndo:** Franco-European Multidisciplinary Endometriosis Institute
**IMMPACT:** Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
**MRI:** magnetic resonance imagery
**NRS:** numerical rating scale
**RCT:** randomized controlled trial
**VR:** virtual reality

---

*Edited by R Kukafka; submitted 13.05.22; peer-reviewed by W Ceron, X Xiong, ER Khalilian, E Rolla, BJ Nievas-Soriano; comments to author 14.07.22; revised version received 26.07.22; accepted 11.08.22; published 21.09.22.*

*Please cite as:*
*Merlot B, Dispersyn G, Husson Z, Chanavaz-Lacheray I, Dennis T, Greco-Vuilloud J, Fougère M, Potvin S, Cotty-Eslous M, Roman H, Marchand S*
*Pain Reduction With an Immersive Digital Therapeutic Tool in Women Living With Endometriosis-Related Pelvic Pain: Randomized Controlled Trial*
*J Med Internet Res 2022;24(9):e39531*
*URL: https://www.jmir.org/2022/9/e39531*
*doi:10.2196/39531*
*PMID:36129733*

Original Paper

# Social Media News Use and COVID-19 Misinformation Engagement: Survey Study

Saifuddin Ahmed[1], MSc, PhD; Muhammad Ehab Rasul[2], MA

[1]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

[2]Department of Communication, University of California Davis, Davis, CA, United States

**Corresponding Author:**
Saifuddin Ahmed, MSc, PhD
Wee Kim Wee School of Communication and Information
Nanyang Technological University
31 Nanyang Link, #03-43
Singapore, 637718
Singapore
Phone: 65 63168910
Email: sahmed@ntu.edu.sg

## Abstract

**Background:** Social media is widely used as a source of news and information regarding COVID-19. However, the abundance of misinformation on social media platforms has raised concerns regarding the spreading infodemic. Accordingly, many have questioned the utility and impact of social media news use on users' engagement with (mis)information.

**Objective:** This study offers a conceptual framework for how social media news use influences COVID-19 misinformation engagement. More specifically, we examined how news consumption on social media leads to COVID-19 misinformation sharing by inducing belief in such misinformation. We further explored if the effects of social media news use on COVID-19 misinformation engagement depend on individual differences in cognition and personality traits.

**Methods:** We used data from an online survey panel administered by a survey agency (Qualtrics) in Singapore. The survey was conducted in March 2022, and 500 respondents answered the survey. All participants were older than 21 years and provided consent before taking part in the study. We used linear regression, mediation, and moderated mediation analyses to explore the proposed relationships between social media news use, cognitive ability, personality traits, and COVID-19 misinformation belief and sharing intentions.

**Results:** The results suggested that those who frequently used social media for news consumption were more likely to believe COVID-19 misinformation and share it on social media. Further probing the mechanism suggested that social media news use translated into sharing intent via the perceived accuracy of misinformation. Simply put, social media news users shared COVID-19 misinformation because they believed it to be accurate. We also found that those with high levels of extraversion than those with low levels were more likely to perceive the misinformation to be accurate and share it. Those with high levels of neuroticism and openness than those with low levels were also likely to perceive the misinformation to be accurate. Finally, it was observed that personality traits did not significantly influence misinformation sharing at higher levels of cognitive ability, but low cognitive users largely drove misinformation sharing across personality traits.

**Conclusions:** The reliance on social media platforms for news consumption during the COVID-19 pandemic has amplified, with dire consequences for misinformation sharing. This study shows that increased social media news consumption is associated with believing and sharing COVID-19 misinformation, with low cognitive users being the most vulnerable. We offer recommendations to newsmakers, social media moderators, and policymakers toward efforts in limiting COVID-19 misinformation propagation and safeguarding citizens.

***(J Med Internet Res 2022;24(9):e38944)*** doi:10.2196/38944

**KEYWORDS**

COVID-19; misinformation; personality; cognitive ability; social media; Singapore

## Introduction

The emergence of the novel coronavirus or SARS-CoV-2 wreaked havoc across the world. One of the consequences of the resulting pandemic was an unprecedented reliance on social media platforms, as public health agencies and governments turned to social media as a tool for news dissemination [1,2]. Therefore, social media platforms provided individuals with quick access to credible information while also allowing them to share their opinions and attitudes toward the pandemic [3,4]. Individuals also used social media to cope with additional stresses, stay-at-home orders, and remote work environments, while displaying what scholars consider to be signs of social media addiction [5]. The deluge of illegitimate, anecdotal, and emotional content created a perfect storm for the emergence of the COVID-19 infodemic involving the "undisciplined spread of information" [6,7], which swept through social media and was rife with rumors, disinformation, and conspiracy theories [8].

The effects of misinformation related to a destructive pandemic, such as COVID-19, are severe because false beliefs are difficult to correct [9], especially among those with low cognitive ability [10]. Moreover, although misinformation on social media is not a novel issue, it has become a key cause of concern owing to its pernicious impact on infectious disease management and public compliance with health protocols such as mask wearing [11]. For instance, researchers have found that misinformation on social media may lead to lower trust in public health authorities and the effectiveness of mitigation protocols [12]. In addition, studies have found that misinformation on social media fuels vaccine hesitancy through engagement with antivaccine beliefs [13,14]. The unabated spread of misinformation online, along with its severe impact on negative attitudes toward science and compliance toward public health protocols, has prompted scientific investigation into the sharing and engagement of misinformation related to COVID-19. However, the susceptibility to misinformation on social media varies from person to person. Scholars have argued that some individuals are more vulnerable to misinformation on social media than others [15]. While prior work has explored the relationship between general social media use and misinformation vulnerability, our concern is on the growing proportion of social media users who rely heavily on it for news updates. We argue that general social media use is distinct from social media news use. General social media use is a broad term that can encompass many activities (ie, liking posts, sharing posts, commenting, watching videos, etc). On the other hand, social media news use is a narrower operationalization of social media news consumption that can aid researchers in examining which specific aspects of social media use impact misinformation engagement. Indeed, a growing body of literature has found social media news use to be positively associated with the spread of conspiracy theories and misinformation [16]. Other studies have found social media news use to play a role in the sharing of deepfakes, a form of misinformation [17,18]. The post-COVID climate is likely to have a magnification effect on these relationships, as more and more people are increasingly turning to social media for news use. Therefore, we anticipate that social media news use offers a nuanced understanding of social media effects with potentially more serious consequences than previously understood.

The technological features of social media platforms play a role in the spread of misinformation on social media. For instance, algorithms that curate the social media feed seek to maximize engagement through prior behavior and clickbait [19]. Consequently, this may result in repeated exposure to misinformation and increase individual engagement with false information related to COVID-19. Existing research has established that repeated exposure to misinformation reinforces and increases trust in false beliefs [20]. Repeated exposure causes individuals to be more susceptible to misinformation through the illusory truth effect, which posits that repeated claims are seen as more truthful than nonrepeated claims [21]. Consequently, it may lead to poor discernment of truthful information and a lack of careful reasoning [22]. In addition to increased susceptibility toward misinformation, individuals may share false beliefs with others in their networks. As such, we hypothesize the following: *H1*, social media news use will be positively associated with (1) perceived accuracy and (2) sharing intentions of COVID-19 misinformation; and *H2*, the relationship between social media news use and sharing intention will be mediated by the perceived accuracy of COVID-19 misinformation.

Scholars have explored the individual-level differences in how people react to misinformation. For instance, prior research has found that political ideology, particularly conservativism, is a predictor and motivator of belief in misinformation [22-24] and of sharing COVID-19 misinformation–related conspiracy theories on social media [25]. Another study suggested that this may be because the heightened levels of anxiety among Republicans led them to trust and share misinformation related to COVID-19 on social media through partisan motivated reasoning and selective sharing [26]. However, the bipartisan American context may not apply to other contexts of social media use and misinformation sharing worldwide. Subsequently, we consider other intergroup differences related to reasoning and rationality that have been explored in the literature. For instance, a recent study found that people with higher levels of analytical thinking were less likely to believe and share COVID-19 misinformation on social media [27]. Likewise, researchers have pointed out that individuals with high cognitive ability are less vulnerable to misinformation on social media [28]. When an individual is exposed to misinformation, increased deliberation and controlled thinking to process this information can lead to more accurate detection of fake news. This can occur due to motivated *system 2* reasoning, a part of the dual-process theory, which argues that analytical thinking can override an individual's intuitive and automatic response to information [22]. Hence, individuals who have high cognitive ability are more likely to deliberate carefully and be skeptical of misinformation [18]. Therefore, we propose the following hypothesis: *H3*, cognitive ability will be negatively associated with (1) perceived accuracy and (2) sharing intentions of COVID-19 misinformation.

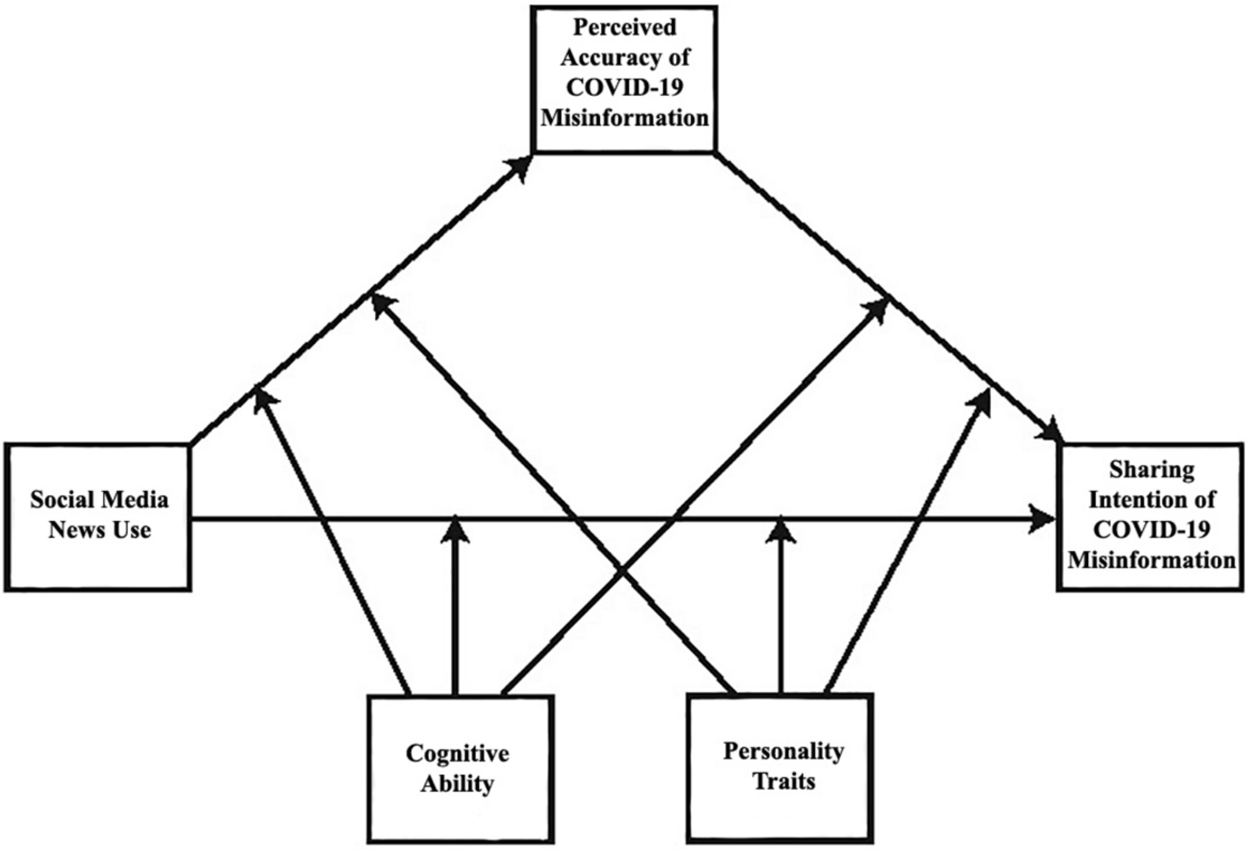
We build on these findings and also consider whether personality traits may illustrate further individual differences in misinformation sharing behavior during the COVID-19 pandemic. Some scholars have suggested that personality traits influence misinformation engagement [29]. We refer to the 5-factor model of personality, which encompasses predispositions to everyday experiences and decision-making through the use of lay adjectives [30]. People's scores of the *big five* personality traits remain relatively stable throughout their lives [31], and scholars have reported replicable relationships of personality traits with different facets of everyday decision-making, including how individuals engage with information, with contrasting findings regarding the role of all personality traits in better news discernment [32,33]. While individual traits do play a role in understanding how individuals engage with information, an overall disposition to manifest "extreme variants" of traits has been linked to compulsive behavior [34], and psychopathy and personality disorders [35]. In the COVID-19 context, personality traits, such as neuroticism, are found to drive beliefs in COVID-19 misinformation and conspiracy theories [36]. We argue that personality traits can offer interesting insights into misinformation engagement in the COVID-19 context, both in terms of themselves and as general indicators of populations with compulsive behavioral tendencies or personality disorders. Furthermore, given the existing literature on the relationship between social media use and personality traits [37,38], there is a need for further research that explores the role of personality traits in COVID-19 misinformation engagement on social media. Hence, we propose the following set of research questions: *RQ1*, "How are personality traits associated with (1) perceived accuracy and (2) sharing intentions of COVID-19 misinformation?" and *RQ2*, "How do personality traits and cognitive ability moderate the mediated relationship between social media news use and sharing intention of COVID-19 misinformation through perceived accuracy?"

To summarize, while the literature on COVID-19 misinformation is growing, there are several gaps that require attention. First, existing studies do not discern between general social media use and news consumption behavior. Social media news consumption behavior is key to understanding misinformation related to COVID-19, as individuals are repeatedly exposed to false information embedded in news stories, particularly from far-right sources [39]. Moreover, social media news use has been linked to COVID-19 vaccine hesitancy, which is a key issue for policy makers worldwide [40]. Second, while rare investigations have focused on the role of personality traits in COVID-19 misinformation engagement on social media [36,41], it has not been studied in conjunction with cognition or cognitive ability, another important factor related to individual engagement with misinformation [42]. Prior research has argued that people with certain personality traits and cognitive ability may engage with misinformation. Specifically, recent studies have found that those with lower cognitive ability are more likely to share false information [17,18]. However, research in this area needs to be expanded upon by scholars. This is also essential because to unravel the psychology of misinformation engagement and to devise counter strategies, we need to consider the individual differences in both personalities and cognition. Lastly, the vast majority of the current literature focuses on Western democracies [8,40,43,44] and largely ignores Asian contexts other than China [45,46]. The Asian population makes up a large portion of social media traffic globally and may be exposed to large amounts of misinformation related to COVID-19, and in turn, this population may share and believe in the false information. For example, according to Statista [47], the estimated number of social media users in Singapore in 2020 was 5.18 million, and this number is expected to increase to 5.68 million by 2025. Moreover, around 83% of Singaporeans seek news online. With such a large number of individuals seeking news online, they may encounter large amounts of false or misleading information. Indeed, approximately 60% of Singaporeans have reported encountering fake news on social media [48]. As such, individuals in Singapore may have been potentially exposed to a large amount of misinformation during COVID-19. This is problematic as Asian countries, such as Singapore, have experienced some of the worst COVID-19 outbreaks. Thus, more attention toward social media misinformation engagement related to COVID-19 is required in Asian contexts.

In addressing the existing research gaps, this study offers a conceptual framework that explains how social media news use influences COVID-19 misinformation engagement. Precisely, we argue that social media news use leads to COVID-19 misinformation sharing through the induction of belief in misinformation (a mediated relationship). Further, we argue that these effects are dependent on the cognition and personality traits of social media users (Figure 1 illustrates the conceptual framework). By identifying individuals who may be more vulnerable to engagement with COVID-19 misinformation, scientists and policy makers can develop strategies to mitigate the harmful effects of false information and encourage compliance with preventative measures. We focus on cognitive ability and personality traits because they have been associated with misinformation engagement in existing literature [18,32]. Moreover, cognitive ability and personality traits are 2 important psychological mechanisms that can influence behavior. Therefore, both personality traits and cognitive ability can provide important insights into COVID-19 misinformation engagement on social media. However, it is also possible that personality traits and cognitive ability affect each other and jointly impact COVID-19 misinformation engagement. In fact, the findings of a recent study suggest that personality traits and cognitive ability interact and affect political misinformation engagement [49]. Specifically, the study found that low cognitive individuals with certain personality traits, such as neuroticism and openness, were more susceptible to engaging with political misinformation. Accordingly, we extend this body of literature by assessing the impact of personality traits and cognitive ability on misinformation engagement in the context of COVID-19.

In order to extend current research, this study relies on survey data from Singapore for several reasons. First, Singapore is an important country in Asia with a diverse population. Second, although a small country, Singapore is one of the most densely populated countries in Asia and the world [50]. Third, according to a recent report by the Reuters Institute, 53% of Singaporeans relied on social media for news during COVID-19 [51], which may have resulted in increased exposure to misinformation related to COVID-19. Lastly, Singapore's diverse population bridges many Asian cultures owing to a large number of migrant workers present in the country. We used a quota sampling strategy based on population demographics, which can allow for increased generalizability of findings focused on how social media news may lead to COVID-19 misinformation sharing through the induction of beliefs in misinformation. Moreover, it allows for an examination of how personality traits and cognitive ability moderate the relationships mentioned above. This study ultimately meaningfully contributes to the large body of literature focused on COVID-19 misinformation sharing and belief on social media in an understudied context.

**Figure 1.** Conceptual framework of the relationship among social media news use, and perceived accuracy and sharing intention of COVID-19 misinformation, with cognitive ability and personality traits as moderators.



## Methods

### Recruitment

The respondents in the study were recruited through an online panel administered by Qualtrics. Qualtrics is an online survey platform that maintains a panel of several million US residents who have volunteered to take part in online surveys. Respondents who complete a survey are compensated by Qualtrics. Qualtrics uses a quota sampling technique to identify and match participants with the study's requirement, with the aim to recruit a sample that closely matches the demographic distribution of the census. We used a similar approach to match the sample to population parameters focusing on age and gender. Such techniques have been used previously to ensure that the findings generalize to the larger population [8,17,18,52].

The survey was conducted in March 2022, and out of 1726 respondents who landed on the survey page, 500 respondents answered the survey (28.97% response rate). The study included Singaporean residents older than 21 years. We focused on respondents 21 years or above since this is considered the legal adult age in Singapore.

### Procedure

After providing consent to participate in this study, the participants first answered questions related to their demographic characteristics, media use habits, cognitive ability, and personality traits. Next, the participants proceeded to the (misinformation) evaluation task. They were informed that they would be presented with a few trending news headlines on social media related to COVID-19. Their task involved carefully reading each news headline and answering related questions before moving to the following headline. All 5 viral news headlines presented to the respondents (reported in the Measures section) were false, according to factchecking websites. The participants were not informed that the headlines were not true, as this would have affected the study findings [53].

### Ethics Approval

The Institutional Review Board at Nanyang Technological University approved the study protocol (IRB-2022-097).

### Measures

Perceived accuracy of COVID-19 misinformation was measured by asking the respondents to rate their level of perceived accuracy (1 [not at all accurate] to 5 [extremely accurate]) for the 5 claims in the news headlines. The scale is based on previous research on the perceived accuracy of news/misinformation headlines [54,55]. The participants were asked how accurate are the claims that (1) coconut is effective in reducing COVID-19 symptoms; (2) the pH miracle lifestyle healing program of alkaline diet, exercise, and healing foods can cure COVID-19; (3) COVID vaccines are dangerous and ineffective against the Omicron variant; (4) mRNA COVID-19 vaccinations cause magnetism by introducing graphene oxide into the blood; and (5) there is no evidence of the COVID-19 virus and no one has isolated and sequenced SARS-CoV-2 from any patient sample. The responses to the 5 items were averaged to create an index of the perceived claim accuracy of misinformation regarding COVID-19 (mean 2.01, SD 1.03; α=.91).

Sharing intention of COVID-19 misinformation was measured by asking respondents how likely (1 [extremely likely to share] to 5 [not at all likely to share]) are they to share these news headlines on their social media profiles. While it is acknowledged that these sharing intentions are hypothetical, such approaches have been previously adopted by scholars to measure misinformation sharing [54,56]. Moreover, self-reports of sharing intentions have been found to be strongly associated with attention received by news headlines on social media [57]. The responses to sharing intentions were reverse coded, so a higher value represents greater sharing intention. The responses were then averaged to create an index of the sharing intention of COVID-19 misinformation (mean 1.96, SD 1.08; α=.93).

Social media news use was measured by asking respondents how frequently (1 [never] to 5 [daily]) do they engage in the following: (1) post on their timeline about political or public affairs news; (2) share posts about political or public affairs news; (3) comment on posts about political or public affairs news; (4) read their news feed about political or public affairs news; and (5) read the news feed/timelines of friends about political or public affairs news [58]. The responses to the 5 items were averaged to create an index of social media news use (mean 2.26, SD 0.90; α=.79).

Cognitive ability was measured by the wordsum test. The test includes 10 questions, where participants are provided with a source word (eg, caprice) and their task involves matching the source word with the closest related word from a list of 5 target words; in this case, the 5 words are (1) value, (2) star, (3) grimace, (4) whim, and (5) inducement. The correct responses to the 10 questions were summed to create a scale of cognitive ability (mean 5.48, SD 2.48; α=.76). While the test is vocabulary based, it has high covariance with general intelligence [59,60] and has been frequently used by scholars to investigate the role of cognitive ability in misinformation engagement [17,18,61,62]. The test is also applicable in Singapore since English is the primary language of the educational system and is also the most commonly used verbal language in the country [63].

Personality traits were measured through a total of 10 statements asking respondents to rate their level of agreement (1 [strongly disagree] to 7 [strongly agree]) for the given statements (eg, “I see myself as someone who worries a lot,” “I see myself as someone who is talkative,” and “I see myself as someone who does a thorough job”). The responses were combined to cover 5 different personality traits, including neuroticism (mean 4.41, SD 1.54; α=.88), extraversion (mean 4.11, SD 1.45; α=.80), openness (mean 4.81, SD 1.17; α=.77), agreeableness (mean 5.06, SD 1.08; α=.72), and conscientiousness (mean 5.24, SD 1.06; α=.77).

### Statistical Analysis

We employed ordinary least squares (OLS) regression models to test the effect of social media news use on perceived accuracy and sharing intentions of COVID-19 misinformation. We ran a mediation model to explore the proposed mediation relationship. Finally, we employed a conditional moderated mediation analysis (using the SPSS PROCESS macro v 3.5) [64] to examine the moderating role of cognitive ability and personality traits in the mediation process.

We also controlled for several variables, including demographics, traditional media news use (television, radio, and print news use averaged; 1 [never] to 5 [daily]; mean 2.83, SD 1.20; α=.73), and political interest (1 [not at all interested] to 5 [extremely interested]; mean 2.82, SD 1.06). Traditional media news use and political interest were included as covariates since they have been found to be important factors in sharing misinformation [17,18,29].

Demographics included (1) age (mean 39.23, SD 14.12 years), (2) gender (51% female), (3) education (1 [no formal education] to 7 [doctoral degree]; mean 4.47, SD 0.99; median Bachelor’s degree), (4) household income (1 [less than SGD $1000] to 11 [more than SGD $20,000]; mean 5.07, SD 2.60; median SGD $7000-$8999), and (5) race (76.6% Chinese majority). A currency exchange rate of SGD $1=US $710.98 is applicable.

## Results

In the first step, we ran regression analyses to predict perceived accuracy and sharing intentions. The results of the OLS regression are plotted in Figures 2 and 3 (see Multimedia Appendix 1 for more details). The results suggested that those who were younger (perceived accuracy: B=−0.007, SE=0.003; *P*=.03; sharing intention: B=−0.006, SE=0.003; *P*=.05) and had higher political interest (perceived accuracy: B=0.177, SE=0.042; *P*<.001; sharing intention: B=0.212, SE=0.045; *P*<.001) were likely to both perceive the misinformation to be accurate and share it on social media.

**Figure 2.** Predicting perceived accuracy of COVID-19 misinformation. The plot includes regression coefficients for all variables.

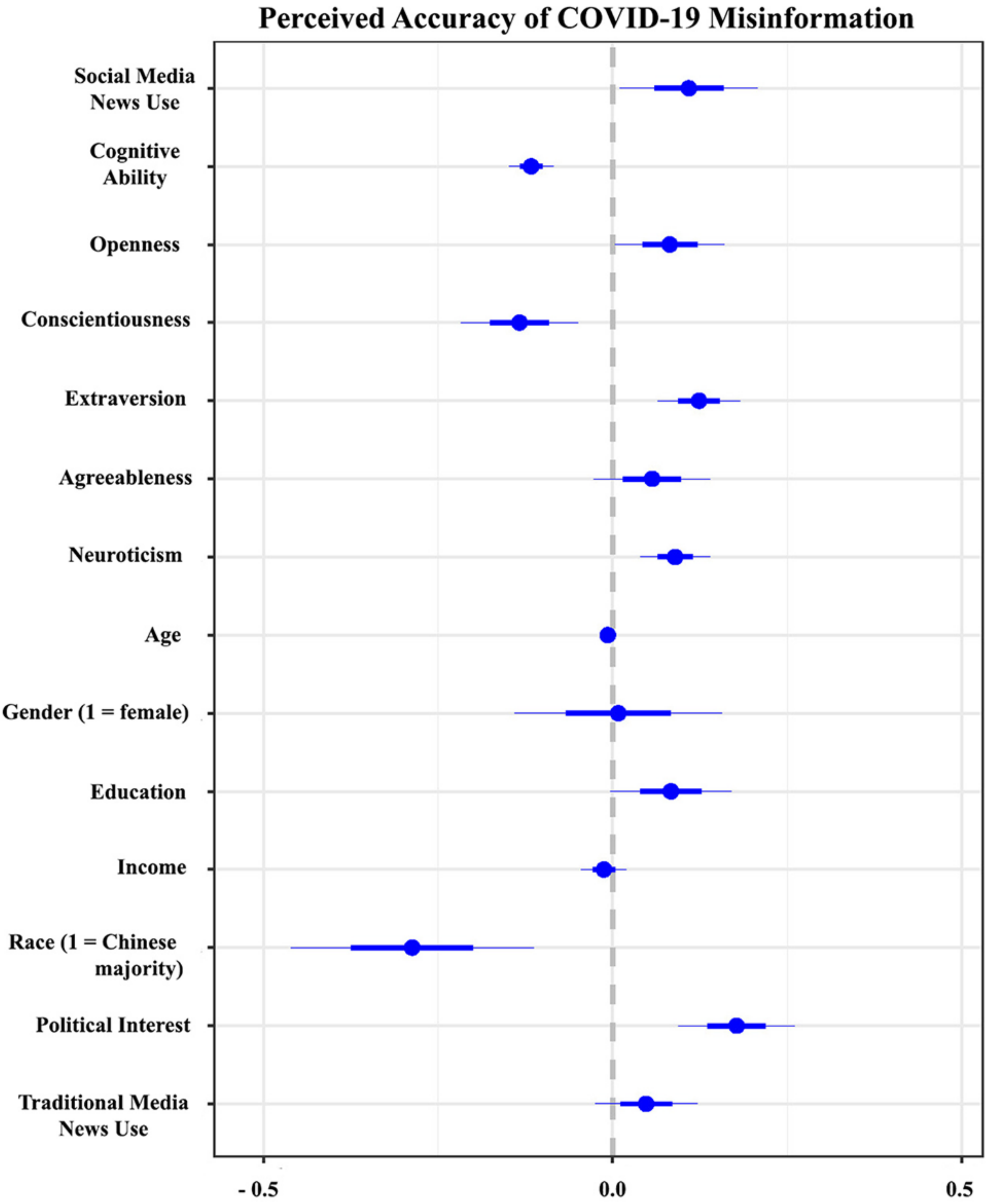
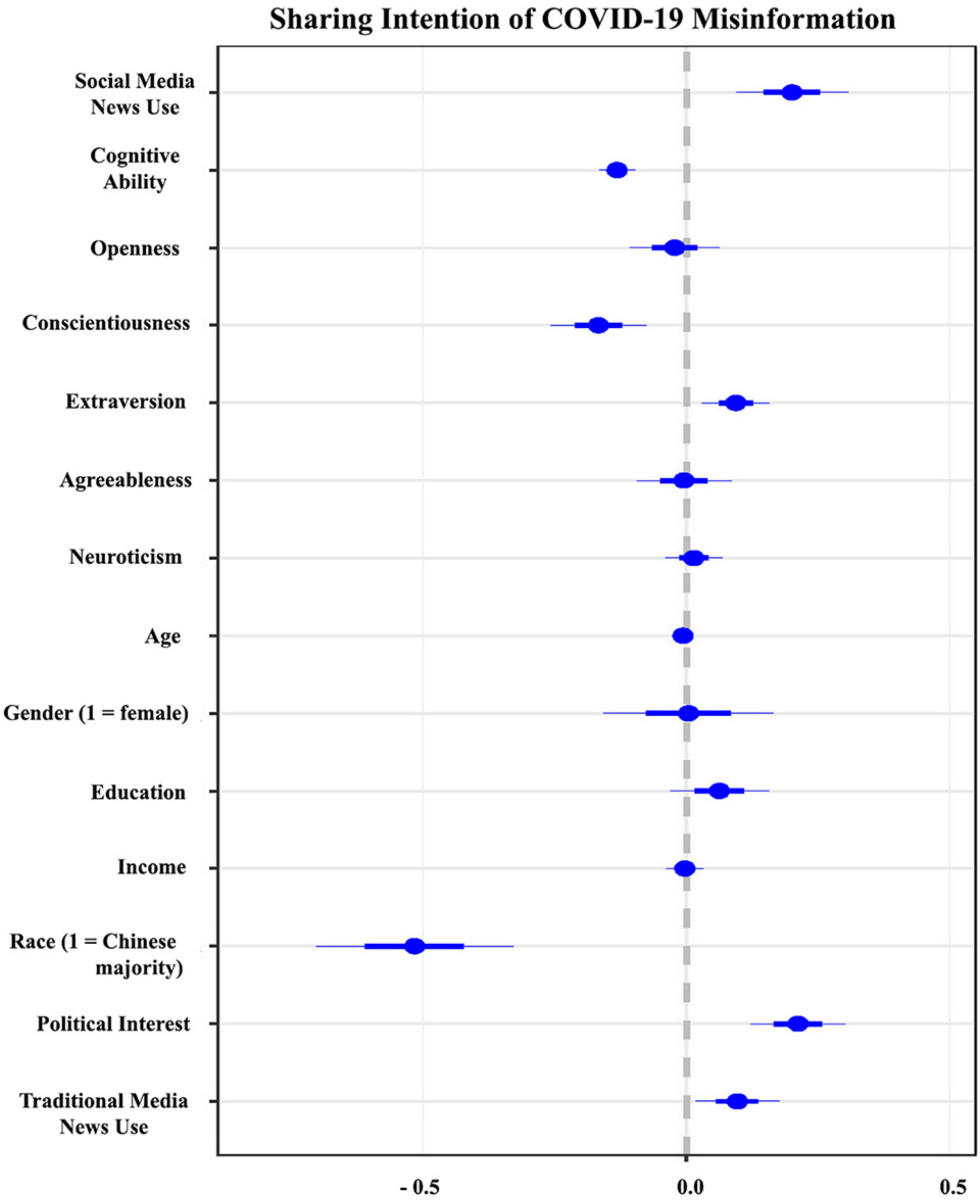


**Figure 3.** Predicting sharing intention of COVID-19 misinformation. The plot includes regression coefficients for all variables.



Among the variables of interest, we observed that those who frequently relied on social media for news consumption were likely to not only perceive the misinformation to be accurate (B=0.109, SE=0.049; *P*=.03) but also share it on social media (B=0.200, SE=0.053; *P*<.001). We also observed that those with high cognitive ability were less likely to perceive the false claims to be true (B=−0.117, SE=0.016; *P*<.001) and share them (B=−0.132, SE=0.017; *P*<.001).

The personality correlates suggested that extraverted individuals were likely to not only perceive the misinformation to be accurate (B=0.124, SE=0.029; *P*<.001) but also show higher sharing intentions (B=0.094, SE=0.032; *P*<.001). Conversely, we observed that conscientious individuals were less likely to perceive the misinformation to be accurate (B=−0.133, SE=0.042; *P*<.001) or share it (B=−0.167, SE=0.045; *P*<.001). In addition, it was also observed that neurotic (B=0.089, SE=0.025; *P*<.001) and open individuals (B=0.082, SE=0.039; *P*=.04) were more likely to perceive the claims to be accurate.

Next, to explore the mechanism of how social media news use induces misinformation sharing intentions through perceived accuracy, we ran mediation analyses using the SPSS PROCESS macro [64], with social media news use as the predictor variable, perceived accuracy as the mediator, and sharing intention as the outcome variable. The bootstrapping method was used to estimate the indirect effects (N=5000).

The results are illustrated in Figure 4. As observed, we found that social media news use (B=0.109, SE=0.049, 95% CI 0.012-0.206) was positively associated with the perceived accuracy of misinformation, which concurrently was positively associated with sharing intentions of misinformation (B=0.697, SE=0.038, 95% CI 0.623-0.771). The direct relationship between social media news use and sharing intentions was also found to be significantly positive (B=0.124, SE=0.041, 95% CI 0.044-0.205).

A formal statistical test of the mediation process suggested that the indirect effects were statistically significant (B=0.076, SE=0.036, 95% CI 0.009-0.147). These results indicated that social media news use translates into sharing intentions of misinformation as individuals perceive this misinformation to be accurate.

Finally, we explored how cognitive ability and personality traits moderated the relationship between social media news use and misinformation sharing intention through perceived accuracy of misinformation. We employed conditional process analyses using the SPSS PROCESS macro for 2 conditional moderators (model 76) [64]. The results of the conditional indirect effects of social media news use on sharing intentions via perceived claim accuracy at different levels (−1 SD, mean, and +1 SD) of cognitive ability and individual personality traits are included in Table 1. While exploring the effects of each personality trait (eg, openness), we used the other 4 components (eg, conscientiousness, extraversion, agreeableness, and neuroticism) as controls in the specific models (see Multimedia Appendix 2 for more details).

The patterns across the 5 personality traits at high levels of cognitive ability suggested that none of the indirect effects (except high cognitive ability and high extraversion) were statistically significant. The general implication is that personality traits do not significantly influence misinformation engagement at higher levels of cognitive ability. On the contrary, it was also found that at lower levels of cognitive ability, individuals who displayed heightened levels of *any* personality trait were more likely to engage with misinformation. In general, this suggests that individuals with low cognitive ability are more susceptible to misinformation if they appear to demonstrate a compulsive outlook toward any of the personality traits.

**Figure 4.** Illustrated mediation of social media news use, perceived accuracy, and sharing intention of COVID-19 misinformation. Estimates are calculated using the SPSS PROCESS macro (Model 4) [64]. The number in parenthesis is the indirect effect with lower limit CI to upper limit CI. Bootstrap resample=5000. Statistical controls include age, gender, education, income, race, political interest, traditional media news use, personality traits, and cognitive ability. ***P*<.01, ****P*<.001.

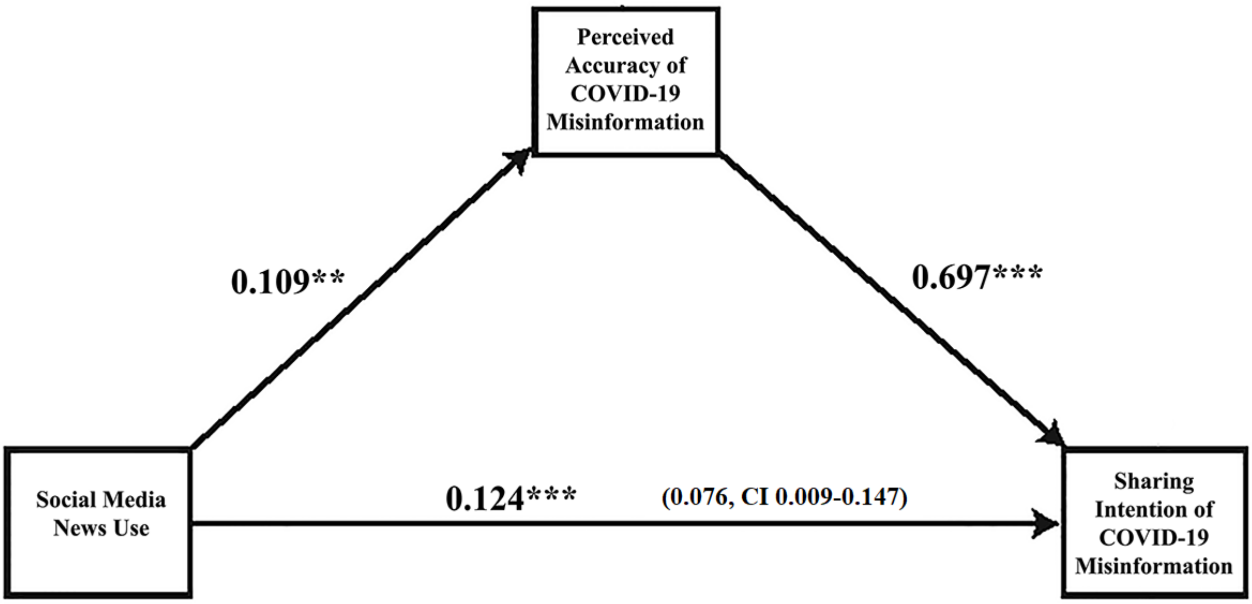


**Table 1.** Conditional indirect effects of social media news use on sharing intentions through perceived accuracy at different levels of cognitive ability and personality traits.

| Cognitive ability level and personality trait level[a] | Effect | Boot SE[b] | LLCI[c] | ULCI[d] |
|---|---|---|---|---|
| Low (−1 SD) cognitive ability and low (−1 SD) openness | 0.12 | 0.07 | −0.01 | 0.25 |
| Low (−1 SD) cognitive ability and mean openness | 0.15[e] | 0.04 | 0.07 | 0.24 |
| Low (−1 SD) cognitive ability and high (+1 SD) openness | 0.18[e] | 0.05 | 0.10 | 0.28 |
| Mean cognitive ability and low (−1 SD) openness | 0.03 | 0.06 | −0.09 | 0.14 |
| Mean cognitive ability and mean openness | 0.07 | 0.04 | −0.01 | 0.15 |
| Mean cognitive ability and high (+1 SD) openness | 0.11[e] | 0.05 | 0.03 | 0.21 |
| High (+1 SD) cognitive ability and low (−1 SD) openness | −0.08 | 0.07 | −0.22 | 0.05 |
| High (+1 SD) cognitive ability and mean openness | −0.03 | 0.05 | −0.14 | 0.07 |
| High (+1 SD) cognitive ability and high (+1 SD) openness | 0.01 | 0.06 | −0.10 | 0.14 |
| Low (−1 SD) cognitive ability and low (−1 SD) conscientiousness | 0.07 | 0.05 | −0.01 | 0.19 |
| Low (−1 SD) cognitive ability and mean conscientiousness | 0.14[e] | 0.04 | 0.07 | 0.23 |
| Low (−1 SD) cognitive ability and high (+1 SD) conscientiousness | 0.24[e] | 0.05 | 0.13 | 0.35 |
| Mean cognitive ability and low (−1 SD) conscientiousness | −0.01 | 0.05 | −0.09 | 0.10 |
| Mean cognitive ability and mean conscientiousness | 0.06 | 0.04 | −0.01 | 0.14 |
| Mean cognitive ability and high (+1 SD) conscientiousness | 0.15[e] | 0.05 | 0.05 | 0.26 |
| High (+1 SD) cognitive ability and low (−1 SD) conscientiousness | −0.11 | 0.06 | −0.22 | 0.00 |
| High (+1 SD) cognitive ability and mean conscientiousness | −0.05 | 0.05 | −0.15 | 0.04 |
| High (+1 SD) cognitive ability and high (+1 SD) conscientiousness | 0.03 | 0.07 | −0.11 | 0.16 |
| Low (−1 SD) cognitive ability and low (−1 SD) extraversion | −0.01 | 0.04 | −0.09 | 0.08 |
| Low (−1 SD) cognitive ability and mean extraversion | 0.09[e] | 0.04 | 0.02 | 0.18 |
| Low (−1 SD) cognitive ability and high (+1 SD) extraversion | 0.23[e] | 0.05 | 0.15 | 0.33 |
| Mean cognitive ability and low (−1 SD) extraversion | −0.06 | 0.04 | −0.13 | 0.01 |
| Mean cognitive ability and mean extraversion | 0.05 | 0.03 | −0.02 | 0.12 |
| Mean cognitive ability and high (+1 SD) extraversion | 0.21[e] | 0.05 | 0.11 | 0.30 |
| High (+1 SD) cognitive ability and low (−1 SD) extraversion | −0.13[e] | 0.05 | −0.23 | −0.05 |
| High (+1 SD) cognitive ability and mean extraversion | −0.01 | 0.05 | −0.11 | 0.09 |
| High (+1 SD) cognitive ability and high (+1 SD) extraversion | 0.16[e] | 0.08 | 0.01 | 0.31 |
| Low (−1 SD) cognitive ability and low (−1 SD) agreeableness | 0.08 | 0.05 | −0.02 | 0.19 |
| Low (−1 SD) cognitive ability and mean agreeableness | 0.14[e] | 0.04 | 0.07 | 0.23 |
| Low (−1 SD) cognitive ability and high (+1 SD) agreeableness | 0.21[e] | 0.05 | 0.12 | 0.31 |
| Mean cognitive ability and low (−1 SD) agreeableness | −0.01 | 0.04 | −0.09 | 0.09 |
| Mean cognitive ability and mean agreeableness | 0.07[e] | 0.04 | 0.00 | 0.14 |
| Mean cognitive ability and high (+1 SD) agreeableness | 0.14[e] | 0.05 | 0.04 | 0.25 |
| High (+1 SD) cognitive ability and low (−1 SD) agreeableness | −0.11 | 0.05 | −0.22 | 0.00 |
| High (+1 SD) cognitive ability and mean agreeableness | −0.03 | 0.05 | −0.13 | 0.07 |
| High (+1 SD) cognitive ability and high (+1 SD) agreeableness | 0.05 | 0.07 | −0.08 | 0.20 |
| Low (−1 SD) cognitive ability and low (−1 SD) neuroticism | 0.11[e] | 0.06 | 0.01 | 0.23 |
| Low (−1 SD) cognitive ability and mean neuroticism | 0.15[e] | 0.04 | 0.07 | 0.23 |

| Cognitive ability level and personality trait level[a] | Effect | Boot SE[b] | LLCI[c] | ULCI[d] |
|---|---|---|---|---|
| Low (−1 SD) cognitive ability and high (+1 SD) neuroticism | 0.19[e] | 0.05 | 0.10 | 0.29 |
| Mean cognitive ability and low (−1 SD) neuroticism | 0.04 | 0.04 | −0.04 | 0.12 |
| Mean cognitive ability and mean neuroticism | 0.08[e] | 0.04 | 0.01 | 0.15 |
| Mean cognitive ability and high (+1 SD) neuroticism | 0.12[e] | 0.06 | 0.02 | 0.25 |
| High (+1 SD) cognitive ability and low (−1 SD) neuroticism | −0.06 | 0.05 | −0.16 | 0.03 |
| High (+1 SD) cognitive ability and mean neuroticism | −0.01 | 0.06 | −0.12 | 0.10 |
| High (+1 SD) cognitive ability and high (+1 SD) neuroticism | 0.04 | 0.08 | −0.11 | 0.21 |

[a] Analyses were performed using the PROCESS macro for SPSS (model 76), applying 5000 bootstrap samples. Statistical controls include age, gender, education, income, race, political trust, political interest, traditional media news use, and personality traits as well as the remaining 4 personality traits.

[b] SE: standard error.

[c] LLCI: lower limit CI.

[d] ULCI: upper limit CI.

[e] Statistically significant effect.

## Discussion

### Principal Findings

Numerous studies have explored public engagement with COVID-19 misinformation on social media [7,11,13,14,36], but not many have explored how news consumption through social media platforms affects such misinformation engagement. This study aimed to explore the mechanism of how social media news use influences believing and sharing COVID-19 misinformation. It also examined individual differences in such engagement through the lens of cognitive and personality factors.

The results of this study are critical in the context of the COVID-19 information environment on social media. A wide majority of the population across societies, including Singapore, rely on social media as a critical source of news and information. More recently, the pandemic made social media platforms relevant as sources of COVID-19 information. Against this background, an association between news consumption via social media and misinformation engagement raises concerns regarding the utility of these platforms as information sources.

The results suggest that frequent reliance on social media for news consumption is associated with increased belief and sharing intentions of COVID-19 misinformation. Moreover, the mediation results indicate that social media news users who believe the COVID-19 misinformation to be accurate are more likely to share it. Furthermore, the study found that personality traits do not significantly influence misinformation sharing at higher levels of cognitive ability. Therefore, high cognitive individuals are less likely to believe or share misinformation irrespective of personality traits. On the contrary, at lower levels of cognitive ability, those with high levels of all personality traits are more vulnerable to COVID-19 misinformation sharing.

The plethora of news and information on social media creates a system of information overload. Given that most users have cognitive biases and do not engage in critical information processing, it is likely that such information overload could explain why increased social media news use is associated with the belief and sharing of COVID-19 misinformation. Indeed, scholars have argued that people fail to think sufficiently before engaging with COVID-19 misinformation [65].

The findings highlight the risks associated with news consumption via social media platforms, but we also found that the observed associations vary by the personality and cognitive ability of individuals. We observed that some personality traits (eg, extraversion and conscientiousness) were associated with sharing intentions but others were not (eg, neuroticism, openness, and agreeableness). However, further probing suggested that the effects of personality traits on sharing intents are driven mainly by low rather than high cognitive social media news users. These results are in line with recent findings where cognitive ability was found to be positively associated with better truth discernment [54,55], weaker belief in false content [17,18,66], and reduced sharing intention of misinformation [56]. In addition, a higher cognitive ability allows individuals to make better risk assessments and filter what information is relevant when placing their trust [67]. Thus, it seems that high cognitive individuals, largely irrespective of their personality traits, are better at processing the false information presented to them and thereby refrain from sharing.

It is important to consider how together with the lack of cognitive ability, a display of high-valued personality traits can further amplify the vulnerability of individuals toward misinformation sharing and engagement. Individuals with low cognitive ability, yet with an extreme tendency to be open to experience, conscientious, extraverted, agreeable, or neurotic, are more likely to fall for and share misinformation. The findings corroborate previous work [33], which has reported on the association of all personality traits with the tendency to believe misinformation; however, we illustrate the importance of cognitive ability as a differentiator in this relationship.

It is important to note here that this relationship is not causal and there are other possible causal relationships that may impact COVID-19 misinformation engagement. For instance, a recent study found evidence of relationships between incidental news exposure on social media, news literacy, and COVID-19 misperceptions [68]. The presence or absence of news literacy could impact how people process news about COVID-19, which, in turn, could dictate how they engage with news related to COVID-19 [69]. Other studies have also found scientific knowledge to play a role in COVID-19 misinformation engagement [65]. Since it has become a politicized issue, ideology could also impact how people engage with COVID-19 misinformation. In fact, political conservatism has been found to be associated with belief in COVID-19 misinformation [15]. Moreover, some social media users are exposed to more misinformation than others, which can lead them to be more susceptible to misinformation through the illusory truth effect. Thus, it could be that individuals who are more susceptible to misinformation also engage with COVID-19 misinformation differently. The unabated spread of misinformation during a destructive global pandemic, such as COVID-19, has raised complex issues and problems. As such, there are a variety of possible factors that can lead to COVID-19 misinformation engagement. Thus, more work is needed to establish causal links with COVID-19 misinformation engagement.

Moreover, the findings presented here are specific to the context of COVID-19 misinformation. While the results on social media news use and cognitive ability are largely consistent with previous literature [17,18], it is to be explored if the patterns for personality traits and the interaction with cognitive ability are consistent across other forms of misinformation.

## Practical Implications

This study has practical implications for those who rely on social media for news. Existing studies have found evidence of widespread misinformation related to COVID-19 on different social media platforms, such as YouTube, Twitter, Instagram, and Reddit [6]. Our results add an additional layer of concern for social media platforms when it comes to combatting COVID-19 misinformation. Some studies have found accuracy nudges to be successful in correcting COVID-19 misinformation as a result of news exposure [65]. However, more work is required in this area to fully understand and develop strategies to fight COVID-19 misinformation and its pernicious effects in contexts other than the United States. While our findings represent active social media news users, they also have implications for those who do not primarily use social media for news consumption. For example, a recent survey found that most online users are exposed to news about COVID-19 on social media even when they are on social media for different purposes [65]. Such incidental exposure to COVID-19 information could also expose these individuals to misinformation, thereby furthering the adverse consequences. In fact, a recent study found that incidental news exposure was related to COVID-19 misperceptions [68].

Additionally, previous studies have found that the older population is more vulnerable to believing and sharing misinformation [20,70]. However, we found that younger respondents were more likely to believe and share COVID-19 misinformation. These results are in line with studies that have found similar patterns [71,72]. For example, a study involving samples from 5 countries found that older individuals were less susceptible to misinformation about COVID-19 [15]. Younger respondents may likely lack the necessary skills and abilities to discern COVID-19 misinformation despite high literacy rates, since it is specialized health information. Further, our results have implications for individuals with lower levels of cognitive ability. For example, a study found that lower cognitive ability is related to increased susceptibility to deepfakes on social media. Furthermore, scholars have argued that people often fail to think sufficiently about the accuracy of content on social media before they share it [22]. As such, those with low cognitive ability could be more at risk to not only believe in COVID-19 misinformation, but also share it with their networks on social media platforms. Thus, researchers and social media platforms must find ways to counter the widespread misinformation related to COVID-19 to promote compliance with public health protocols.

## Limitations

The results are based on cross-sectional data and limit any causal inferences. While the findings confirm the overall consensus of the impact of social media on misinformation engagement during COVID-19, future scholars should collect longitudinal data to make causal arguments. The findings are based on a single context where social media penetration is high and strict governmental regulations largely control misinformation. Therefore, how these findings would apply to societies with low social media penetration remains unanswered. Next, while our operationalization of social media news use (focusing on political and public affairs news) is consistent with a majority of the literature [58,73], it remains to be seen how the effects of social media news would differ based on other forms of news use (eg, health news). Finally, we focused on one aspect of cognitive ability (through the lens of verbal reasoning). Others may consider comparing the effects of different types of intelligence (eg, fluid vs crystallized) since various forms of cognitive ability may have differential impacts.

## Conclusion

Social media platforms are increasingly being used as news aggregators and primary news sources by citizens worldwide. Individual differences in user behavior can lead to users being less or more vulnerable to misinformation engagement, and individuals with low cognitive ability and compulsive personality traits are at a further disadvantage as compared with others. We recommend that policymakers and social media giants should consider targeted interventions that aim at understanding and checking patterns in everyday behavior that could amplify individual risk of encountering or sharing misinformation. We also recommend experiments with interventions to curb the spread of COVID-19 misinformation.

## Acknowledgments

This work was completed at Nanyang Technological University and the University of California, Davis. The work was supported by Nanyang Technological University's Start Up Grant.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Predicting the perceived accuracy and sharing intention of COVID-19 misinformation.
[DOCX File , 27 KB - jmir_v24i9e38944_app1.docx ]

Multimedia Appendix 2
Conditional direct effects of social media news use on sharing intentions through perceived accuracy at different levels of cognitive ability and personality traits.
[DOCX File , 32 KB - jmir_v24i9e38944_app2.docx ]

## References

1. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]
2. Tsao S, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt Z. What social media told us in the time of COVID-19: a scoping review. The Lancet Digital Health 2021 Mar;3(3):e175-e194. [doi: 10.1016/S2589-7500(20)30315-0]
3. Karami A, Zhu M, Goldschmidt B, Boyajieff H, Najafabadi MM. COVID-19 Vaccine and Social Media in the U.S.: Exploring Emotions and Discussions on Twitter. Vaccines (Basel) 2021 Sep 23;9(10):1059 [FREE Full text] [doi: 10.3390/vaccines9101059] [Medline: 34696167]
4. Tao Y. Who should apologise: Expressing criticism of public figures on Chinese social media in times of COVID-19. Discourse & Society 2021 May 09;32(5):622-638. [doi: 10.1177/09579265211013116]
5. Zhao N, Zhou G. COVID-19 Stress and Addictive Social Media Use (SMU): Mediating Role of Active Use and Social Media Flow. Front Psychiatry 2021;12:635546 [FREE Full text] [doi: 10.3389/fpsyt.2021.635546] [Medline: 33633616]
6. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
7. Ferrara E, Cresci S, Luceri L. Misinformation, manipulation, and abuse on social media in the era of COVID-19. J Comput Soc Sci 2020;3(2):271-277 [FREE Full text] [doi: 10.1007/s42001-020-00094-5] [Medline: 33251373]
8. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
9. Chan M, Jones C, Hall Jamieson K, Albarracín D. Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. Psychol Sci 2017 Nov;28(11):1531-1546 [FREE Full text] [doi: 10.1177/0956797617714579] [Medline: 28895452]
10. De keersmaecker J, Roets A. 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. Intelligence 2017 Nov;65:107-110. [doi: 10.1016/j.intell.2017.10.005]
11. Vraga EK, Bode L. Addressing COVID-19 Misinformation on Social Media Preemptively and Responsively. Emerg Infect Dis 2021 Feb;27(2):396-403 [FREE Full text] [doi: 10.3201/eid2702.203139] [Medline: 33395379]
12. Lovari A. Spreading (Dis)Trust: Covid-19 Misinformation and Government Intervention in Italy. MaC 2020 Jun 25;8(2):458-461. [doi: 10.17645/mac.v8i2.3219]
13. Hernandez R, Hagen L, Walker K, O'Leary H, Lengacher C. The COVID-19 vaccine social media : healthcare providers' missed dose in addressing misinformation and vaccine hesitancy. Hum Vaccin Immunother 2021 Sep 02;17(9):2962-2964 [FREE Full text] [doi: 10.1080/21645515.2021.1912551] [Medline: 33890838]
14. Muric G, Wu Y, Ferrara E. COVID-19 Vaccine Hesitancy on Social Media: Building a Public Twitter Data Set of Antivaccine Content, Vaccine Misinformation, and Conspiracies. JMIR Public Health Surveill 2021 Nov 17;7(11):e30642. [doi: 10.2196/30642] [Medline: 34653016]
15. Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci 2020 Oct;7(10):201199 [FREE Full text] [doi: 10.1098/rsos.201199] [Medline: 33204475]
16. Xiao X, Borah P, Su Y. The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs. Public Underst Sci 2021 Nov;30(8):977-992 [FREE Full text] [doi: 10.1177/0963662521998025] [Medline: 33663279]

17. Ahmed S. Who inadvertently shares deepfakes? Analyzing the role of political interest, cognitive ability, and social network size. Telematics and Informatics 2021 Mar;57:101508. [doi: 10.1016/j.tele.2020.101508]
18. Ahmed S. Disinformation Sharing Thrives with Fear of Missing Out among Low Cognitive News Users: A Cross-national Examination of Intentional Sharing of Deep Fakes. Journal of Broadcasting & Electronic Media 2022 Feb 23;66(1):89-109. [doi: 10.1080/08838151.2022.2034826]
19. Iyengar S, Massey D. Scientific communication in a post-truth society. Proc Natl Acad Sci U S A 2019 Apr 16;116(16):7656-7661 [FREE Full text] [doi: 10.1073/pnas.1805868115] [Medline: 30478050]
20. Pennycook G, Cannon T, Rand D. Prior exposure increases perceived accuracy of fake news. J Exp Psychol Gen 2018 Dec;147(12):1865-1880 [FREE Full text] [doi: 10.1037/xge0000465] [Medline: 30247057]
21. Dechêne A, Stahl C, Hansen J, Wänke M. The truth about the truth: a meta-analytic review of the truth effect. Pers Soc Psychol Rev 2010 May;14(2):238-257 [FREE Full text] [doi: 10.1177/1088868309352251] [Medline: 20023210]
22. Pennycook G, Rand D. The Psychology of Fake News. Trends Cogn Sci 2021 May;25(5):388-402 [FREE Full text] [doi: 10.1016/j.tics.2021.02.007] [Medline: 33736957]
23. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. Science 2019 Jan 25;363(6425):374-378. [doi: 10.1126/science.aau2706] [Medline: 30679368]
24. Rothgerber H, Wilson T, Whaley D, Rosenfeld D, Humphrey M, Moore A, et al. Politicizing the COVID-19 Pandemic: Ideological Differences in Adherence to Social Distancing. PsyArXiv. Preprint posted online April 22, 2020 [FREE Full text] [doi: 10.31234/osf.io/k23cv]
25. Havey N. Partisan public health: how does political ideology influence support for COVID-19 related misinformation? J Comput Soc Sci 2020;3(2):319-342 [FREE Full text] [doi: 10.1007/s42001-020-00089-2] [Medline: 33163686]
26. Freiling I, Krause N, Scheufele D, Brossard D. Believing and sharing misinformation, fact-checks, and accurate information on social media: The role of anxiety during COVID-19. New Media & Society 2021 Apr 22:146144482110114. [doi: 10.1177/14614448211011451]
27. Nurse M, Ross R, Isler O, Van Rooy D. Analytic thinking predicts accuracy ratings and willingness to share COVID-19 misinformation in Australia. Mem Cognit 2022 Feb;50(2):425-434 [FREE Full text] [doi: 10.3758/s13421-021-01219-5] [Medline: 34453286]
28. Tandoc E, Lee J, Chew M, Tan F, Goh Z. Falling for fake news: the role of political bias and cognitive ability. Asian Journal of Communication 2021 Jun 17;31(4):237-253. [doi: 10.1080/01292986.2021.1941149]
29. Halpern D, Valenzuela S, Katz J, Miranda JP. From Belief in Conspiracy Theories to Trust in Others: Which Factors Influence Exposure, Believing and Sharing Fake News. In: Meiselwitz G, editor. Social Computing and Social Media. Design, Human Behavior and Analytics. HCII 2019. Lecture Notes in Computer Science(), vol 11578. Cham: Springer; 2019:217-232.
30. McCrae RR, Costa PT. A five-factor theory of personality. In: Pervin LA, John OP, editors. Handbook of personality: Theory and research. New York, NY: Guilford Press; 1999:139-153.
31. Soto C, John O. Development of big five domains and facets in adulthood: mean-level age trends and broadly versus narrowly acting mechanisms. J Pers 2012 Aug;80(4):881-914 [FREE Full text] [doi: 10.1111/j.1467-6494.2011.00752.x] [Medline: 22091969]
32. Calvillo D, Garcia R, Bertrand K, Mayers T. Personality factors and self-reported political news consumption predict susceptibility to political fake news. Personality and Individual Differences 2021 May;174:110666. [doi: 10.1016/j.paid.2021.110666]
33. Wolverton C, Stevens D. The impact of personality in recognizing disinformation. OIR 2019 Dec 19;44(1):181-191. [doi: 10.1108/OIR-04-2019-0115]
34. Rector N, Hood K, Richter M, Michael Bagby R. Obsessive-compulsive disorder and the five-factor model of personality: distinction and overlap with major depressive disorder. Behaviour Research and Therapy 2002 Oct;40(10):1205-1219. [doi: 10.1016/S0005-7967(02)00024-4]
35. Miller J, Lynam DR, Widiger T, Leukefeld C. Personality disorders as extreme variants of common personality dimensions: can the Five-Factor Model adequately represent psychopathy? J Pers 2001 Apr;69(2):253-276 [FREE Full text] [doi: 10.1111/1467-6494.00144] [Medline: 11339798]
36. Ferreira S, Campos C, Marinho B, Rocha S, Fonseca-Pedrero E, Barbosa Rocha N. What drives beliefs in COVID-19 conspiracy theories? The role of psychotic-like experiences and confinement-related factors. Soc Sci Med 2022 Jan;292:114611 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114611] [Medline: 34864602]
37. Özgüven N, Mucan B. he relationship between personality traits and social media use. Soc Behav Pers 2013 Apr 01;41(3):517-528. [doi: 10.2224/sbp.2013.41.3.517]
38. Lin JS, Lee YI, Jin Y, Gilbreath B. Personality Traits, Motivations, and Emotional Consequences of Social Media Usage. Cyberpsychol Behav Soc Netw 2017 Oct;20(10):615-623. [doi: 10.1089/cyber.2017.0043] [Medline: 29039699]
39. Motta M, Stecula D, Farhart C. How Right-Leaning Media Coverage of COVID-19 Facilitated the Spread of Misinformation in the Early Stages of the Pandemic in the U.S. Can J Pol Sci 2020 May 01;53(2):335-342. [doi: 10.1017/S0008423920000396]

40. Ahmed S, Rasul ME, Cho J. Social Media News Use Induces COVID-19 Vaccine Hesitancy Through Skepticism Regarding Its Efficacy: A Longitudinal Study From the United States. Front Psychol 2022;13:900386 [FREE Full text] [doi: 10.3389/fpsyg.2022.900386] [Medline: 35756213]
41. Vally Z. Chapter 15 - Compliance with health-protective behaviors in relation to COVID-19: The roles of health-related misinformation, perceived vulnerability, and personality traits. In: Moustafa AA, editor. Mental Health Effects of COVID-19. Cambridge, MA: Academic Press; 2021.
42. Wang X, Chao F, Yu G, Zhang K. Factors influencing fake news rebuttal acceptance during the COVID-19 pandemic and the moderating effect of cognitive ability. Comput Human Behav 2022 May;130:107174 [FREE Full text] [doi: 10.1016/j.chb.2021.107174] [Medline: 35002055]
43. Su Y. It doesn't take a village to fall for misinformation: Social media use, discussion heterogeneity preference, worry of the virus, faith in scientists, and COVID-19-related misinformation beliefs. Telematics and Informatics 2021 May;58:101547. [doi: 10.1016/j.tele.2020.101547]
44. Lee J, Choi J, Britt RK. Social Media as Risk-Attenuation and Misinformation-Amplification Station: How Social Media Interaction Affects Misperceptions about COVID-19. Health Commun 2021 Nov 09:1-11. [doi: 10.1080/10410236.2021.1996920] [Medline: 34753361]
45. Leng Y, Zhai Y, Sun S, Wu Y, Selzer J, Strover S, et al. Misinformation During the COVID-19 Outbreak in China: Cultural, Social and Political Entanglements. IEEE Trans. Big Data 2021 Mar 1;7(1):69-80. [doi: 10.1109/tbdata.2021.3055758]
46. Chen K, Luo Y, Hu A, Zhao J, Zhang L. Characteristics of Misinformation Spreading on Social Media During the COVID-19 Outbreak in China: A Descriptive Analysis. RMHP 2021 May;Volume 14:1869-1879. [doi: 10.2147/rmhp.s312327]
47. Number of internet users in Singapore from 2010 to 2020 and a forecast up to 2025. Statista. URL: https://www.statista.com/statistics/553773/number-of-internet-users-singapore/ [accessed 2022-06-25]
48. 6 in 10 people in Singapore have received fake COVID-19 news, likely on social media: Survey. Channel News Asia. URL: https://www.channelnewsasia.com/singapore/fake-covid-19-news-study-ncid-messaging-platforms-whatsapp-673131 [accessed 2022-06-25]
49. Ahmed S, Tan HW. Personality and perspicacity: Role of personality traits and cognitive ability in political misinformation discernment and sharing behavior. Personality and Individual Differences 2022 Oct;196:111747. [doi: 10.1016/j.paid.2022.111747]
50. Countries with the highest population density worldwide in 2019. Statista. URL: https://www.statista.com/statistics/264683/top-fifty-countries-with-the-highest-population-density [accessed 2022-06-25]
51. Singapore. Reuters Institute for the Study of Journalism. URL: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021/singapore [accessed 2022-06-14]
52. Kim A, Dennis A. Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media. MISQ 2019 Jan 1;43(3):1025-1039. [doi: 10.25300/MISQ/2019/15188]
53. Valenzuela S, Halpern D, Katz J, Miranda J. The Paradox of Participation Versus Misinformation: Social Media, Political Engagement, and the Spread of Misinformation. Digital Journalism 2019 Jun 12;7(6):802-823. [doi: 10.1080/21670811.2019.1623701]
54. Pennycook G, Rand D. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition 2019 Jul;188:39-50 [FREE Full text] [doi: 10.1016/j.cognition.2018.06.011] [Medline: 29935897]
55. Pennycook G, Bear A, Collins E, Rand D. The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. Management Science 2020 Nov;66(11):4944-4957. [doi: 10.1287/mnsc.2019.3478]
56. Ahmed S. Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes. Personality and Individual Differences 2021 Nov;182:111074. [doi: 10.1016/j.paid.2021.111074]
57. Mosleh M, Pennycook G, Rand D. Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. PLoS One 2020;15(2):e0228882. [doi: 10.1371/journal.pone.0228882] [Medline: 32040539]
58. Ahmed S, Chen VHH, Jaidka K, Hooi R, Chib A. Social media use and anti-immigrant attitudes: evidence from a survey and automated linguistic analysis of Facebook posts. Asian Journal of Communication 2021 May 25;31(4):276-298. [doi: 10.1080/01292986.2021.1929358]
59. Caplan B, Miller S. Intelligence makes people think like economists: Evidence from the General Social Survey. Intelligence 2010 Nov;38(6):636-647. [doi: 10.1016/j.intell.2010.09.005]
60. Littrell S, Fugelsang J, Risko E. Overconfidently underthinking: narcissism negatively predicts cognitive reflection. Thinking & Reasoning 2019 Jul 05;26(3):352-380. [doi: 10.1080/13546783.2019.1633404]
61. Greene C, Nash R, Murphy G. Misremembering Brexit: partisan bias and individual predictors of false memories for fake news stories among Brexit voters. Memory 2021 May;29(5):587-604. [doi: 10.1080/09658211.2021.1923754] [Medline: 33971789]
62. Murphy G, Loftus E, Grady R, Levine L, Greene C. False Memories for Fake News During Ireland's Abortion Referendum. Psychol Sci 2019 Oct;30(10):1449-1459. [doi: 10.1177/0956797619864887] [Medline: 31432746]

63. General Household Survey 2015. Government of Singapore. URL: https://www.singstat.gov.sg/publications/ghs/ghs2015 [accessed 2022-04-21]
64. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: Guilford Press; 2017.
65. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. Psychol Sci 2020 Jul;31(7):770-780 [FREE Full text] [doi: 10.1177/0956797620939054] [Medline: 32603243]
66. Ståhl T, van Prooijen J. Epistemic rationality: Skepticism toward unfounded beliefs requires sufficient cognitive ability and motivation to be rational. Personality and Individual Differences 2018 Feb;122:155-163. [doi: 10.1016/j.paid.2017.10.026]
67. Zmerli S, van der Meer TWG. Handbook on Political Trust. Cheltenham, United Kingdom: Edward Elgar Publishing; 2017.
68. Borah P, Su Y, Xiao X, Lai Lee D. Incidental news exposure and COVID-19 misperceptions: A moderated-mediation model. Computers in Human Behavior 2022 Apr;129:107173. [doi: 10.1016/j.chb.2021.107173]
69. Vraga E, Tully M, Bode L. Empowering Users to Respond to Misinformation about Covid-19. MaC 2020 Jun 25;8(2):475-479. [doi: 10.17645/mac.v8i2.3200]
70. Guess A, Nagler J, Tucker J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Sci Adv 2019 Jan;5(1):eaau4586 [FREE Full text] [doi: 10.1126/sciadv.aau4586] [Medline: 30662946]
71. Baptista J, Correia E, Gradim A, Piñeiro-Naval V. The Influence of Political Ideology on Fake News Belief: The Portuguese Case. Publications 2021 May 27;9(2):23. [doi: 10.3390/publications9020023]
72. Chadwick A, Vaccari C. News sharing on UK social media: Misinformation, disinformation, and correction. Loughborough University. 2019. URL: https://www.lboro.ac.uk/media/media/subjects/communication-media-studies/downloads/chadwick-vaccari-o3c-1-news-sharing-on-uk-social-media-1.pdf [accessed 2022-04-21]
73. Gil de Zúñiga H, Weeks B, Ardèvol-Abreu A. Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics. J Comput-Mediat Comm 2017 Apr 11;22(3):105-123. [doi: 10.1111/jcc4.12185]

## Abbreviations

**OLS:** ordinary least squares

---

*Edited by M Gisondi, M Strehlow; submitted 22.04.22; peer-reviewed by W Pan, A Chang; comments to author 17.05.22; revised version received 24.07.22; accepted 14.08.22; published 20.09.22.*

*Please cite as:*
*Ahmed S, Rasul ME*
*Social Media News Use and COVID-19 Misinformation Engagement: Survey Study*
*J Med Internet Res 2022;24(9):e38944*
*URL: https://www.jmir.org/2022/9/e38944*
*doi:10.2196/38944*
*PMID:36067414*

Original Paper

# Co-Development of a Web Application (COVID-19 Social Site) for Long-Term Care Workers ("Something for Us"): User-Centered Design and Participatory Research Study

Catherine H Saunders[1], MPH, PhD; Ailyn Sierpe[1], MSc; Gabrielle Stevens[1], PhD; Glyn Elwyn[1], MSc, MD, PhD; Matthew Cantrell[2], NHA; Jaclyn Engel[1], BA; Melissa Gonzalez[1]; Martha Hayward[3], BA; Joellen Huebner[3], MPH; Lisa Johnson[1], MBA; Alejandro Jimenez[4]; Nancy Ruth Little[5], MPH, EdD; Corinne McKenna[4], BA; Manu Onteeru[1]; May Oo Khine[1]; Jacqueline Pogue[1], MPH; José Luis Salinas Vargas[1]; Peter Schmidt[6], PhD; Rachael Thomeer[1], BA; CONFIDENT Study Long-Term Care Partners[1]; Marie-Anne Durand[1,7], MSc, MPhil, PhD

[1]The Dartmouth Institute for Health Policy & Clinical Practice, Dartmouth College, Lebanon, NH, United States
[2]National Association of Health Care Assistants, Carl Junction, MO, United States
[3]Institute for Healthcare Improvement, Boston, MA, United States
[4]Yellow Lion Media, New York City, NY, United States
[5]Department of Public Health, East Carolina University, Greenville, NC, United States
[6]Department of Neurology, NYU Grossman School of Medicine, New York City, NY, United States
[7]Centre for Epidemiology and Research in Population Health, Université de Toulouse, Toulouse, France

**Corresponding Author:**
Catherine H Saunders, MPH, PhD
The Dartmouth Institute for Health Policy & Clinical Practice
Dartmouth College
1 Medical Center Dr
Lebanon, NH, 03766
United States
Phone: 1 (603) 653 0800
Email: catherine.hylas.saunders@dartmouth.edu

## Abstract

**Background:** Improving confidence in and uptake of COVID-19 vaccines and boosters among long-term care workers (LTCWs) is a crucial public health goal, given their role in the care of elderly people and people at risk. While difficult to reach with workplace communication interventions, most LTCWs regularly use social media and smartphones. Various social media interventions have improved attitudes and uptake for other vaccines and hold promise for the LTCW population.

**Objective:** We aimed to develop a curated social web application (interactive website) to increase COVID-19 vaccine confidence (a 3-arm randomized trial is underway).

**Methods:** Following user-centric design and participatory research approaches, we undertook the following 3 steps: (1) content identification, (2) platform development, and (3) community building. A LTCW and stakeholder advisory group provided iterative input. For content identification (step 1), we identified topics of concern about COVID-19 vaccines via desktop research (published literature, public opinion polls, and social media monitoring), refined by interviewing and polling LTCWs. We also conducted a national online panel survey. We curated and fact-checked posts from popular social media platforms that addressed the identified concerns. During platform development (step 2), we solicited preferences for design and functionality via interviews and user experience testing with LTCWs. We also identified best practices for online community building (step 3).

**Results:** In the interviews (n=9), we identified 3 themes: (1) LTCWs are proud of their work but feel undervalued; (2) LTCWs have varying levels of trust in COVID-19–related information; and (3) LTCWs would welcome a curated COVID-19 resource that is easy to understand and use-"something for us". Through desktop research, LTCW interviews, and our national online panel survey (n=592) we found that participants are interested in information about COVID-19 in general, vaccine benefits, vaccine risks, and vaccine development. Content identification resulted in 434 posts addressing these topic areas, with 209 uploaded to the final web application. Our LTCW poll (n=8) revealed preferences for personal stories and video content. The platform we developed is an accessible WordPress-based social media web application, refined through formal (n=3) and informal user experience testing. Users can sort posts by topic or subtopic and react to or comment on posts. To build an online community, we recruited 3 LTCW “community ambassadors” and instructed them to encourage discussion, acknowledge concerns, and offer factual information on COVID-19 vaccines. We also set “community standards” for the web application.

**Conclusions:** An iterative, user-centric, participatory approach led to the launch of an accessible social media web application with curated content for COVID-19 vaccines targeting LTCWs in the United States. Through our trial, we will determine if this approach successfully improves vaccine confidence. If so, a similar social media resource could be used to develop curated social media interventions in other populations and with other public health goals.

***(J Med Internet Res 2022;24(9):e38359)*** doi:10.2196/38359

**KEYWORDS**

COVID-19; vaccine hesitancy; long-term care; social media; web application; website; intervention development, information and communications technology

## Introduction

COVID-19 vaccination rates among long-term care workers (LTCWs) vary across the United States, partly due to a patchwork of legal challenges to a Centers for Medicare and Medicaid Services vaccination mandate [1]. Full vaccination rates range from 70% to 99%, with booster rates trailing from 17% to 56% [2].

Long-term care settings have been major outbreak sites throughout the COVID-19 pandemic, leading to illness and death among vulnerable residents and staff. Researchers of the Centers for Disease Control and Prevention estimate that more than 2300 LTCWs and 151,000 residents have died from COVID-19. COVID-19 outbreaks threaten LTCWs themselves and their often underserved communities [3]. More than half of LTCWs are from disadvantaged socioeconomic, racial, or ethnic groups [4]. LTCWs from underserved communities and those with lower educational attainment are less likely to be vaccinated for COVID-19 than those from advantaged backgrounds [5]. Long-term care staff shortages predate the pandemic, but recent data from the Bureau of Labor Statistics show a further loss of nearly 400,000 LTCWs from 2020 to 2021 [6,7].

Because LTCWs care for the frailest elderly, increasing vaccine confidence and uptake in this population while simultaneously supporting and retaining the LTCW workforce is a critical public health need.

Although improving vaccine confidence and uptake among LTCWs is challenging, social media is a promising potential solution. Social media–based interventions can suit marginalized groups and LTCWs already relying on social media for information (personal communication, Matthew Cantrell, April 2021)[8]. Randomized trials of social media interventions have shown some success, largely in improving attitudes about other vaccines [9-12]. When combined with selected information from medical experts and communication between participants, they have also improved vaccine uptake [10,13]. As far as we know, no social media web applications target LTCWs and address their questions and concerns about COVID-19 vaccines. It is unknown whether this type of intervention would improve confidence or uptake of COVID-19 vaccines and boosters.

We aimed to fill this gap by developing a curated social media web application for LTCWs with low confidence in COVID-19 vaccines.

## Methods

### Study Design

We conducted a user-centered design and participatory research study to develop, qualitatively assess (usability and acceptability), and deploy a social media web application (interactive website) called the COVID-19 Social Site [14,15]. We curated and customized the site for LTCWs with low COVID-19 vaccine confidence within a broader randomized clinical trial (ClinicalTrials.gov, NCT05168800), funded by the Patient-Centered Outcomes Research Institute (COVID-2021C2-13181).

We reported results using the Consolidated Criteria for Reporting Qualitative Research (CORE-Q) and the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [16,17].

### Participatory Approach

Our National Association of Health Care Assistants (NAHCA) partners were critical to web application development. Their deep expertise in long-term care informed development from conception to launch.

We recruited 10 LTCW partners from diverse backgrounds and positions as part of a stakeholder advisory group, which met regularly. Our partners were instrumental in designing and developing the web app, and providing iterative feedback and advice throughout all stages of the project. Their feedback is captured throughout this manuscript (Multimedia Appendix 1).

### Theoretical Framework and Context

We conceived this intervention within a broader comparative effectiveness trial with 2 interventions guided by the theoretical framework by Peretti-Watel et al for vaccine hesitancy, which considers vaccine hesitancy a decision-making process, not a static state [18,19]. Therefore, interventions that increase knowledge in the right context could also increase vaccine confidence and uptake [20]. Emerging evidence suggests that multi-component dialogue-based interventions can be effective, particularly when context and hesitancy drivers are taken into account. Developers must also tailor content, format, and delivery to specific audiences [20].

## Ethics Approval

Dartmouth College's Committee for the Protection of Human Subjects approved this study (STUDY00032340).

## Step 1. Content Identification

### *Desktop Research*

We first identified the common questions and concerns associated with low COVID-19 vaccine confidence by reviewing information from the published literature, public opinion polls, and social media. Given the evolving pandemic, we had a flexible search strategy (Multimedia Appendix 2).

We developed a dynamic list of questions and concerns by consulting with our team (including LTCW partners and other stakeholders) and cross-checking with existing resources [21]. We grouped the questions and concerns into top-level topics. We refined topic wording with plain language principles [22].

### *LTCW Stakeholder Consultation*

We shared our top-level topics with our LTCW stakeholder advisors via semistructured interviews. We have provided details on the interview population, recruitment, procedures, and analysis (Multimedia Appendix 3 [22,23]), and the interview guide (Multimedia Appendix 4)[23,24].

We polled our LTCW partners to assess the content mix they wanted on the web application, including questions about the source (platform and creator), type, quantity, and tone of posts. Our questionnaire is presented in Multimedia Appendix 5.

### *National Online Panel Survey*

We deployed a Qualtrics (Seattle, WA) survey (Multimedia Appendix 6) using members of an existing panel to gain insights from a model population on the perceived importance of different COVID-19 vaccine-related information. Multimedia Appendix 7 [25-32] provides further details about survey development, sampling, recruitment, and analysis [4,25,26].

### *Content Curation*

We sourced material from popular social media web applications according to the preferences identified by LTCWs via interviews and a poll. We plan to continue sourcing content (Multimedia Appendix 2) throughout the life of the site.

### *Content Processing and Fact-Checking*

The content team identified social media posts of interest and logged them along with basic details (eg, date posted, the platform of origin, and engagement metrics). We used a category-based system with hashtags.

We designed a fact-checking process (Multimedia Appendix 8) in consultation with our broader advisory group, including LTCW partners and other stakeholders. The study team also reviewed each post to confirm appropriateness in light of content mix preferences.

## Step 2. Platform Development

### *LTCW Stakeholder Consultation*

We presented our initial concept for the web application to our LTCW stakeholders, soliciting information about the desired look, feel, and functionality iteratively and during the semistructured interviews mentioned in Step 1. We have provided details on the methods in Multimedia Appendix 3.

### *Web Application Wireframes and Initial Build*

We shared our initial concept informed by stakeholder insight with a web design and development company. They advised on the functionality, and look and feel of the web application. The web team developed initial wireframes and the preliminary site on WordPress, primarily using Blade, Javascript, Hypertext Preprocessor (PHP), and Sassy Cascading Style Sheets (SCSS) scripting languages. We iteratively modified and tested the site with the study team, LTCW partners, and other stakeholders.

### *User Experience Testing*

After testing and modifying initial wireframes with LTCW partners and other stakeholders, we conducted one-on-one user experience (UX) and user interface (UI) testing sessions with LTCWs who were naive to the study and its goals. We also solicited UX and UI feedback from LTCW partners and other stakeholders during meetings. Multimedia Appendix 9 [26, 33-37] provides more details on our testing approach, including our affinity mapping analysis [33-38].

## Step 3. Community Building

### *LTCW Stakeholder Consultation*

During interviews and stakeholder meetings, we asked LTCWs how they wanted the web application to operate and what kind of moderation or guidance we should use. We reviewed the community standards of popular Facebook groups focusing on COVID-19 vaccine discussion. Further details on our approach to community building are provided in Multimedia Appendix 10 [21].

# Results

## Participatory Approach

LTCWs were integral to every stage of this web application development project. Our LTCW partners noted:

> *Certified nursing assistants are often overlooked and dismissed. We help individuals every day, but it is an honor to be given the opportunity to help so many people on such a large scale.* [LTCW stakeholder]
>
> *To have my opinion considered and appreciated gives me great satisfaction and encourages me to continue empowering my profession.* [LTCW stakeholder]

## Step 1. Content Identification

### *Desktop Research*

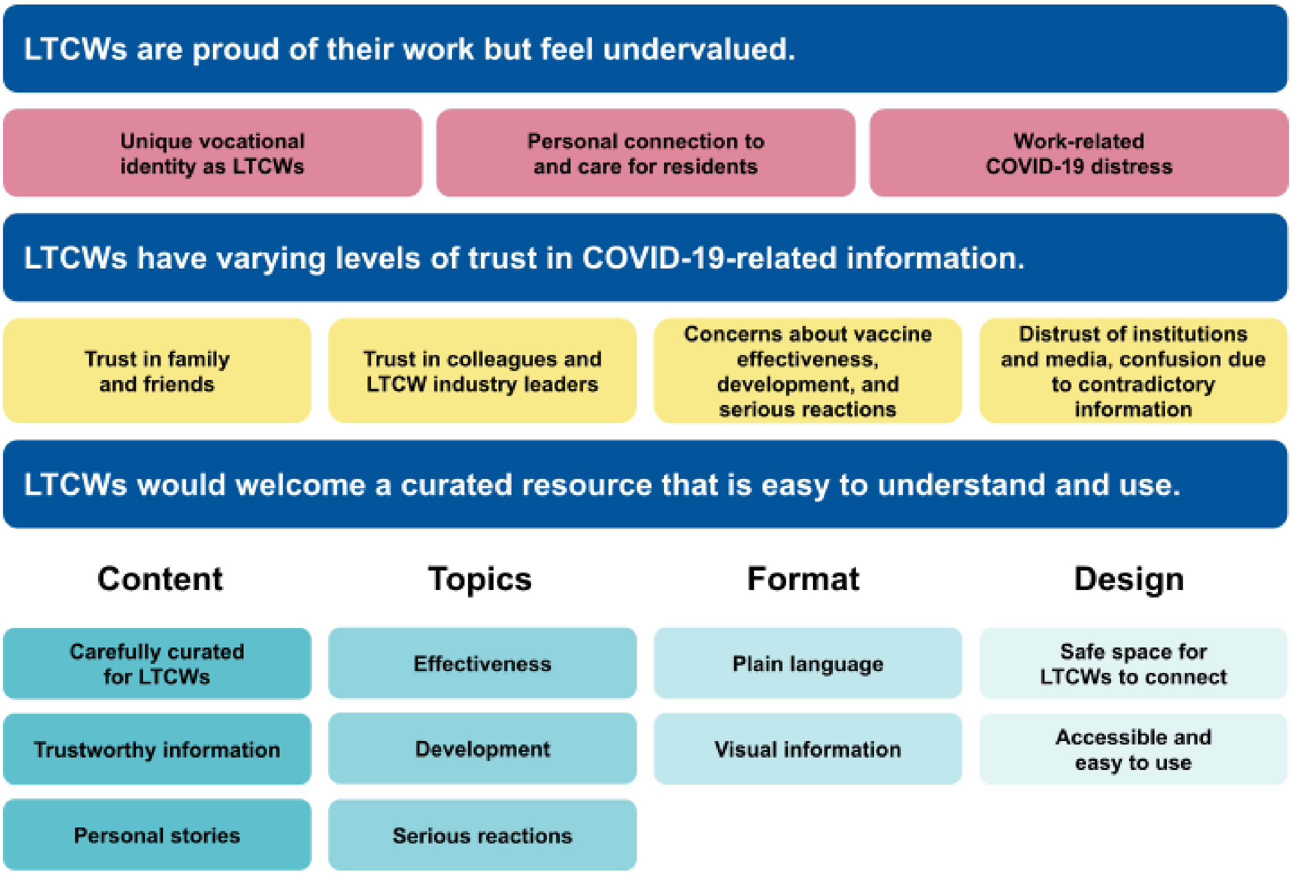
We grouped questions and concerns about COVID-19, and COVID-19 vaccines and boosters into the following 5 general topics: access and process, benefits and efficacy, side effects and harm, development process, and the pandemic overall.

### *LTCW Stakeholder Consultation*

We interviewed 9 LTCWs (5 certified nursing assistants, 1 food services worker, 1 activities director, 1 maintenance worker, and 1 other direct care worker) working in skilled nursing facilities, home care, and hospice care across various regions in the United States. The LTCWs interviewed included 5 females and 4 males. Four participants were White, 2 were Asian, 2 were Black, and 1 was biracial. All LTCWs indicated English as their preferred language, except for 1 who selected Cantonese. Interviews lasted approximately an hour and occurred over Zoom (Zoom Video Communications) between September 16 and November 5, 2021.

We found 3 major themes (Figure 1; Multimedia Appendix 1).

**Figure 1.** Long-term care worker interview themes. LTCW: long-term care workers.



LTCWs are proud of their work but feel undervalued.

Unique vocational identity as LTCWs

Personal connection to and care for residents

Work-related COVID-19 distress

LTCWs have varying levels of trust in COVID-19-related information.

Trust in family and friends

Trust in colleagues and LTCW industry leaders

Concerns about vaccine effectiveness, development, and serious reactions

Distrust of institutions and media, confusion due to contradictory information

LTCWs would welcome a curated resource that is easy to understand and use.

| Content | Topics | Format | Design |
|---|---|---|---|
| Carefully curated for LTCWs | Effectiveness | Plain language | Safe space for LTCWs to connect |
| Trustworthy information | Development | Visual information | Accessible and easy to use |
| Personal stories | Serious reactions | | |

## Theme 1. LTCWs are Proud of Their Work and Yet Feel Undervalued

All interviewees expressed pride and unique vocational identity as LTCWs. They spoke impassionedly, with phrases like

> *Where my heart is.* [Participant #6]
>
> *I don’t call it my passion, I don’t call it my calling; it’s my ministry.* [Participant #4]

A minority of participants noted frustration that LTCWs are generally unrecognized in favor of nursing or other health care workers.

LTCWs spontaneously cited their care for and connection to long-term care residents as central to their professional identities.

> *I love that it’s such a sense of family when we’re there.* [Participant #2]

This sense of responsibility was a powerful motivator for the LTCWs who decided to get COVID-19 vaccines.

> *In my line of work, I work with the demographic most at risk [so] I have a moral responsibility to other people.* [Participant #2]

The web of professional and emotional connections magnified the loss and trauma LTCWs experienced during the pandemic, with most noting death and illness among their residents, colleagues, families, and communities. One participant made the following statement:

> *Seeing these people that I work with – and I love – [...] in a very quick amount of time go from a healthy senior to gone was very devastating.* [Participant #2]

Another participant recounted their experience of near hospitalization due to COVID-19, which prompted them to get vaccinated.

## Theme 2. LTCWs Have Varying Levels of Trust in Information About COVID-19, and COVID-19 Vaccines and Boosters

During the COVID-19 pandemic, LTCWs received information from various sources, including their employers, families and friends, the government, and news and social media. Participants were more likely to trust their families and friends about COVID-19, and its vaccines and boosters than other sources. Sometimes these influences resulted in vaccination, and other times they increased hesitancy. One participant made the following statement:

> *[My husband] is the one that is hesitant [about] the vaccine, and I am with him. He is the one that influenced me not to get it right now.* [Participant #3]

LTCWs noted their unique access to up-to-date information, citing the medical directors at their facilities and industry leaders as strong influences. Interviewees viewed most other mainstream information sources, including the government and the news media, as unreliable. One participant made the following statement:

> *I am not a scientist or a doctor. It's very hard to find information that's accurate.* [Participant #2]

In part because of their mixed information sources and partly due to the contradictory nature of official COVID-19 messages, LTCWs have outstanding questions about COVID-19 vaccines. One participant made the following statement:

> *Even the CDC and the FDA [...] still have different voices about the vaccine.* [Participant #6]

Another encapsulated LTCWs' concerns with the following statement:

> *It's [...] still a trial. They do not know 100% the consequences, the side effects, long-term side effects, [...] the ingredients.* [Participant #9]

Of particular concern were vaccine effectiveness, the vaccine development process, and potential harms, including long-term effects (fertility and unknown future problems) and serious reactions (myocarditis and blood clots). One participant recalled a colleague who made the following statement:

> *Her only reason was because she didn't want to end up on a commercial 10 years from now that said, "Have you developed this, this, and this and this because of the [...] COVID vaccine that was forced on you 10 years ago? You may be liable for a lawsuit [sic]"* [Participant #7]

### Theme 3. LTCWs Would Welcome a Carefully Curated, Easy-to-Understand and Use COVID-19 Resource

Outstanding questions about COVID-19 and vaccines meant most LTCWs wanted a dedicated place where they could find trustworthy information.

> *Targeted at who you're trying to get* [Participant #4]

Knowing they could trust the information was critical.

> *Just not knowing how to find out, not knowing how to research, or not knowing how to look into the resources. I think it's really important that be made easily and readily available to people.* [Participant #2]

LTCW participants said the best way to get curated and trustworthy information was through personal stories and plain clear communication. One participant made the following statement:

> *Having actual stories to all of this kind of stuff is going to probably be key. People need to connect with other people.* [Participant #5]

Clear simple communication is paramount, especially given the busy nature of LTCWs' lives. One interviewee said they needed the following:

> *Solid information, but in a simpler way, like easy to understand.* [Participant #1]

Other participants preferred visual information.

> *I think pictures speak louder than words to a lot of people, especially with social media.* [Participant #7]

### Content Topics

Concerning the web application content itself, our LTCW interviewees appreciated the 5 proposed topics but suggested we eliminate the access and process topic. It was clear to LTCWs where and how they could get vaccinated. Additionally, they emphasized the importance of including content about COVID-19 vaccine effectiveness, the development process, and the potential short- and long-term risks. They also worked with us to refine the topic names for clarity (Multimedia Appendix 2). Finally, our stakeholder partners expressed interest in light, non-COVID, LTCW-related topics, tapping into their professional identity.

### Content Mix

When asked about the desired characteristics of the COVID-19 Social Site's content, we found that the LTCW stakeholders wanted overall diversity in post format. Infographics and text-based content were most and least favored, respectively. The group strongly preferred content from Facebook and YouTube, with TikTok as the least preferred. They favored posts from certified nursing assistants, scientists, and laypeople and disfavored content from journalists. The LTCW stakeholders slightly preferred serious content over lighthearted content. Finally, they slightly preferred evergreen content to content covering new developments. Additional details are provided in Multimedia Appendix 5.

## *National Online Panel Survey*

A total of 592 participants comprised the final survey sample. Participant flow and characteristics are detailed in Multimedia Appendix 7.

A large proportion of participants rated each information topic as at least "a little" important (Table 1; Multimedia Appendix 6). How effectively the vaccines protect people from the virus was most frequently the most important category among respondents. Participants least frequently selected vaccine benefits as the most important category. The other 4 categories were closely grouped. No new topics meaningfully different from the existing topics emerged from open-text responses with sufficient frequency to include. Additional results are provided in Multimedia Appendix 7.

**Table 1.** Ranked importance of COVID-19 topics in a national online panel survey.

| Topic rank | Ratings of "a little important" or more per topic category (averaged across items) | | Topic category most frequently of top importance per respondent | |
|---|---|---|---|---|
| | Category | Percentage | Category | Percentage |
| 1 | How well the vaccines work | 94% | How well the vaccines work | 46% |
| 2 | Overall COVID-19 impact | 92% | Overall COVID-19 impact | 37% |
| 3 | Vaccine benefits | 91% | Vaccine creation | 35% |
| 4 | Vaccine creation | 90% | Long-term vaccine problems | 34% |
| 5 | Short-term vaccine side effects | 89% | Short-term vaccine side effects | 33% |
| 6 | Long-term vaccine problems | 89% | Vaccine benefits | 24% |

### *Final Content*

By launch, we had identified 434 content items for the web application, with 209 items ultimately uploaded to the live site (Figure 2). Post characteristics varied (Table 2). Most posts (n=133) were videos, and the remainder were images (n=65) or text-based posts (n=11). The content came largely from Instagram (n=56), TikTok (n=54), and YouTube (n=51), with fewer posts from Facebook, Twitter, and Reddit.

Most posts (n=125) addressed the benefits of getting vaccinated, such as preventing death and illness. About a third (n=75) of the content covered COVID-19 as a disease and the severity of the pandemic. Other posts were about vaccine development (n=39) and the potential risks associated with vaccination (n=35), including side effects and heart problems. Some posts addressed multiple categories. Finally, a minority (n=25) of posts were not specifically related to these topic categories. These posts were largely health care workers' personal stories, discussions of how vaccine misinformation works in general, or memes from certified nursing assistants or LTCWs.

We worked to reflect the LTCWs' content mix preferences and the national survey result topics in our final content set, although some preferences competed, including the preference for video but suspicion of sources on TikTok. Given that the concerns about TikTok were related to its perceived untrustworthiness, we worked to rigorously fact-check each post to mitigate this concern.

**Figure 2.** Content identification, fact-checking, and screening flow diagram.

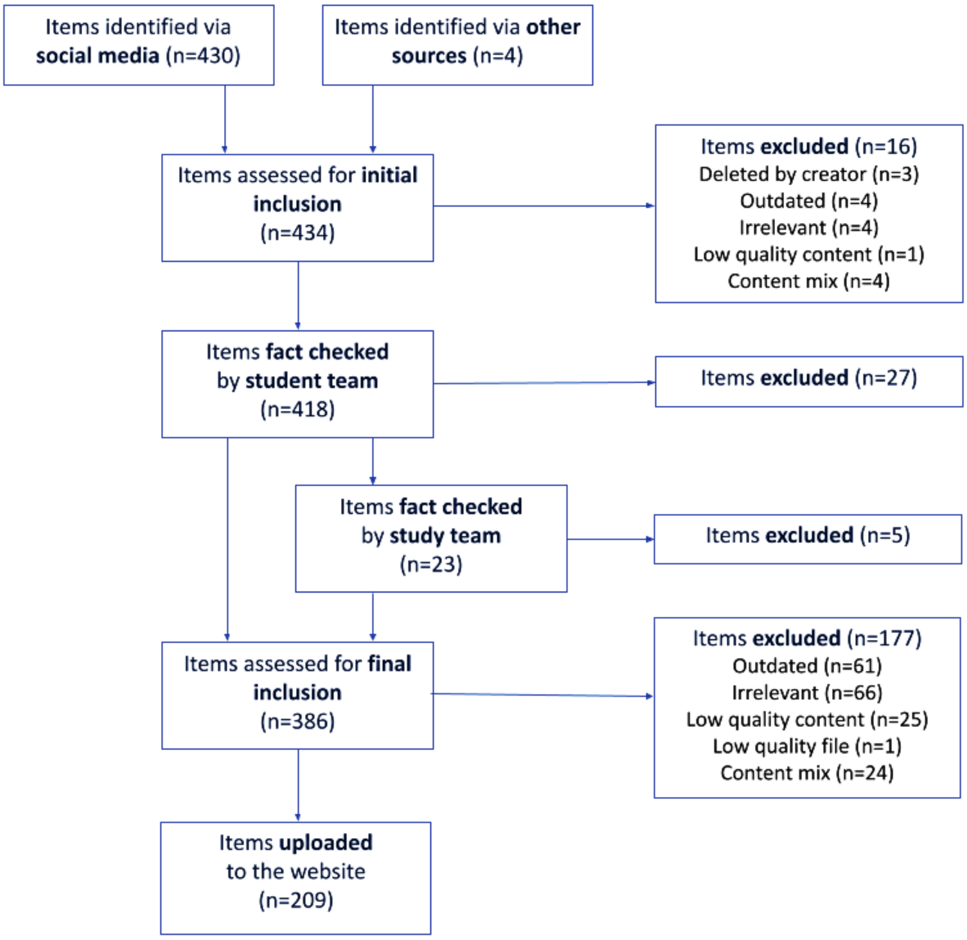


**Table 2.** Final content characteristics.

| Characteristic | Value (N=209), n (%) |
|---|---|
| **Media type** | |
| Video | 133 (63.6) |
| Image | 65 (31.1) |
| Text | 11 (5.3) |
| **Media source** | |
| Instagram | 56 (26.8) |
| TikTok | 54 (25.8) |
| YouTube | 51 (24.4) |
| Facebook | 32 (15.3) |
| Twitter | 8 (3.8) |
| Reddit | 6 (2.9) |
| Other | 2 (1.0) |
| **Creator role** | |
| Medical expert | 87 (41.6) |
| Journalist | 27 (12.9) |
| Healthcare organization | 22 (10.5) |
| Government | 20 (9.6) |
| Layperson | 20 (9.6) |
| Long-term care worker or certified nursing assistant | 18 (8.6) |
| University or education organization | 14 (6.7) |
| Study team | 1 (0.5) |
| **Topics**[a] | |
| Vaccine benefits | 125 (59.8) |
| About COVID-19 | 75 (35.9) |
| Vaccine creation | 39 (18.7) |
| Vaccine risks | 35 (16.7) |
| General | 25 (12.0) |

[a] Individual content items may address multiple topics.

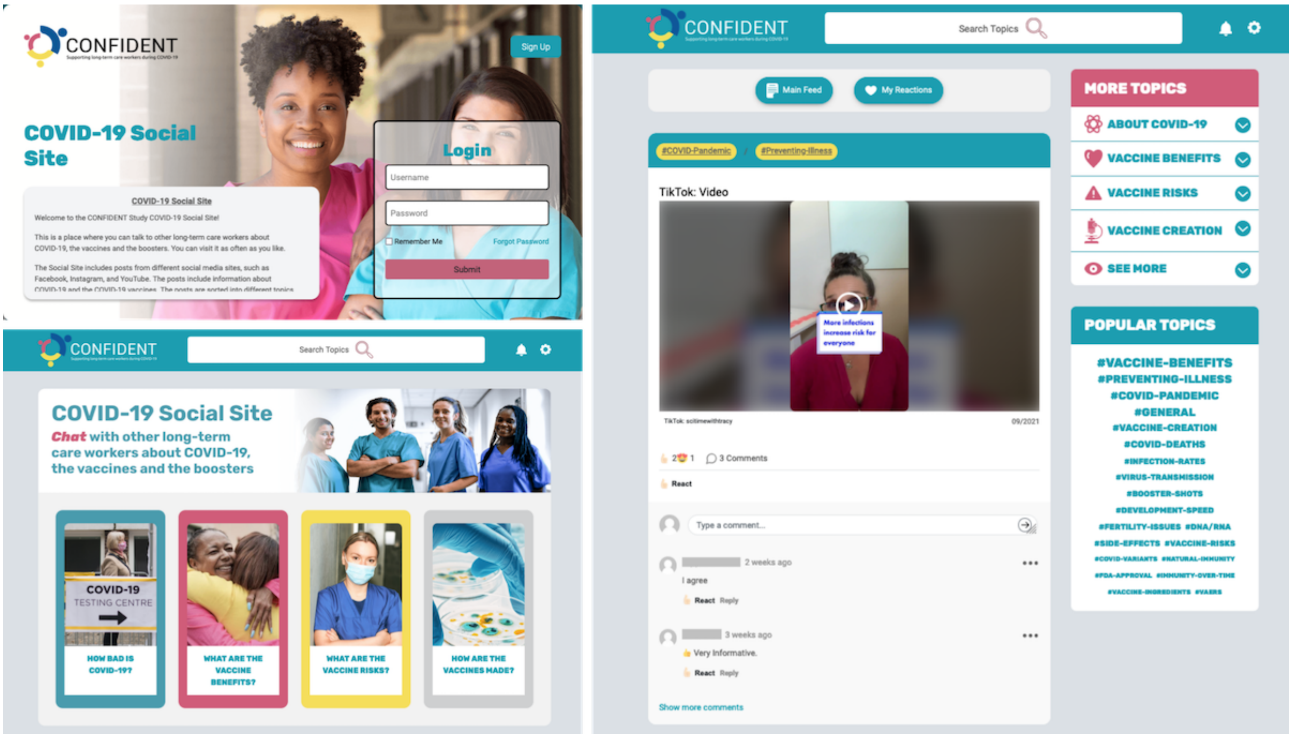
## Step 2. Platform Development

We created a WordPress-based social media web application called the COVID-19 Social Site (Figure 3). We specifically curated it for LTCWs. It featured an infinite scroll feed with information about each post (title, date, and source). LTCWs could sort posts by topic (level 1) or subtopic (level 2) via an expanding sidebar menu. They could also sort by specific hashtags (level 3).

Users could react to posts with emojis (labeled Like, Love, Haha, Wow, Sad, and Angry). They could also comment on posts and reply or react to other users' comments. A notification bell alerted users to new activity, and we sent them email updates. We optimized the web application for desktop and mobile use. Interactivity was consistent with other popular social media web applications [15].

Additionally, we tracked user interaction across the social web application using Google Analytics. We informed users of the data collection via cookies with a pop-up dialog box that appeared on their first visit to the site. These web analytics services allowed us to determine which topics or posts received the most views, measure overall user engagement, and troubleshoot navigational or technical issues.

**Figure 3.** COVID-19 Social Site login and home page.



COVID-19 Social Site

Welcome to the CONFIDENT Study COVID-19 Social Site!

This is a place where you can talk to other long-term care workers about COVID-19, the vaccines and the boosters. You can visit it as often as you like.

The Social Site includes posts from different social media sites, such as Facebook, Instagram, and YouTube. The posts include information about

[redacted] 2 weeks ago

I agree

[redacted] 3 weeks ago

Very Informative.

### LTCW Stakeholder Consultation

LTCW stakeholder consultation participant characteristics are detailed in Step 1. The subthemes that emerged from our interviews with LTCWs related to platform development were that (1) the web application needed to be accessible and easy to use, and that (2) LTCWs should have a safe space to interact with each other, particularly through comments and likes.

### Accessible and Easy to Use

Although the user habits and levels of digital literacy of the LTCWs and their peers varied, almost all interviewees wanted a clearly laid out web application with simple navigation. We reflected these preferences in the site's final design, prioritizing ease of use.

### A Safe Space for LTCWs to Connect

LTCWs wanted the web application to facilitate engagement with other LTCWs via comments. A few participants expressed concerns about the possibility of interpersonal conflict or vaccine-related misinformation but noted that careful moderation by the study team could mitigate this risk. The LTCWs also wanted to interact with posts and comments with likes, emojis, or similar. Multiple interviewees shared that while browsing established social media platforms like Facebook, they were most likely to view and actively engage with posts that had many reactions and comments.

### UX Testing

We conducted 3 formal UX interviews with LTCWs and various informal UX tests with study team members and stakeholders [26].

We found that users easily recognized the web application as a social media platform and understood the layout. All interviewees agreed that the language and topics featured were valuable. We uncovered various opportunities to decrease user pain points by improving the navigation experience. For instance, we added an instructional video and cues when hovering over the main navigation menus to encourage clicking. Additional details are provided in Multimedia Appendix 9. We did not conduct repeat interviews.

## Step 3. Community Building

### Community Standards and Moderation

Our final community standards (Multimedia Appendix 10 included guidance about not giving or soliciting medical advice, and avoiding harassment, profanity, hate speech, and spamming. They also encouraged participants to maintain user privacy by not sharing information about others outside the web application.

### Community Ambassadors

Given the emphasis on a safe space specifically for LTCWs, we decided LTCWs should play a visible role in the web application instead of the research staff, who were present behind the scenes but not visible. Through stakeholder connections, we recruited 3 students as "community ambassadors." All 3 had long-term care experience and were training to become nursing home administrators.

The ambassadors were special users and community members, empowered by the study team. They used the site freely, commenting and reacting based on their views and positionality as LTCWs.

We asked the community ambassadors to report any comments of concern for individual review by moderators. This approach allowed the site to function as a partially self-moderating community, as regular users could also flag comments for review.

In addition to monitoring the site for community standard violations, we instructed the community ambassadors to encourage participant discussion by actively engaging with other users' comments. We also asked the community ambassadors to respond to questions and concerns using information relating to COVID-19, the vaccines, and the boosters vetted by the study team (Multimedia Appendix 11).

Additional details on the community standards and the community ambassadors are outlined in Multimedia Appendix 10.

# Discussion

## Principal Findings

We successfully co-developed a novel social media web application featuring curated content specifically tailored to LTCWs with low vaccine confidence. LTCWs are a difficult-to-reach population who we were able to engage through participatory research and user-centered design. Although LTCWs report feeling overlooked professionally, they enthusiastically participated as partners in our project. Testing of the web application in a randomized trial is ongoing.

Through formative semistructured interviews, we found that LTCWs are proud of their work but feel undervalued, particularly in light of the COVID-19 pandemic. They often distrust official sources of information, including the government and popular media. Due to this combination of professional identity and distrust, they welcomed a social media platform specifically customized to them that is trustworthy and easy to understand and use.

To our knowledge, this is the first co-developed web application using curated content from social media to influence attitudes and behaviors about COVID-19 vaccines and boosters. It is also the first web application of its kind specifically for LTCWs, a critically important but understudied population.

## Limitations

We developed the web application in response to the COVID-19 pandemic, a rapidly evolving and urgent public health crisis. Web application development, testing, and launch occurred on an expedited timeline. Under more favorable conditions, we would have conducted more extensive user testing before the trial launch.

A key limitation of our qualitative stakeholder work is that we conducted most of our formative interviews with LTCWs who were already engaged in the project and vaccinated. This sample may have affected the responses. Additionally, LTCWs who agreed to become partners in a vaccine confidence project may be meaningfully different from other LTCWs. While most of our stakeholders were vaccinated, they still had questions and concerns about COVID-19 vaccines and boosters.

Although we designed our national online panel survey to include individuals who were demographically representative of LTCWs, the information preferences of LTCWs may be inherently different from those of the survey participants. Additionally, our content mix poll included a small sample, limiting its representativeness.

Our social media web application, siloed in its own space online, is materially different from most other social media as it exists separately from the rest of the information ecosystem. This separation was necessary to isolate the effects of our intervention and prevent contamination in a larger randomized trial.

## Comparison With Prior Work

Through our user-centered design and participatory research approach, our intervention evolved to reflect other successful interventions in this space more closely, namely a social media web application trialed by Glanz et al that improved vaccine uptake among children of participating parents [10]. This intervention featured vetted information from the study team and ways for participants to interact with each other and the researchers [10].

Other research teams have successfully developed patient-facing communication interventions using participatory research methods [14]. Moderation of the intervention's content by stakeholders is a participatory research approach we have never implemented before. Participatory research is promising and increasingly popular, and although evidence of efficacy or effectiveness is limited, the evidence base is growing [39-42]. We look forward to contributing to this body of knowledge with our randomized controlled trial results. While meaningful stakeholder engagement in all stages of the intervention development and project may be challenging in a condensed timeline, in the context of a pandemic, the benefits highlighted in the context of our study far outweighed the constraints.

Concerning the qualitative experiences of LTCWs overall, our findings that they feel underappreciated professionally are consistent with the findings in the literature, including research that has emerged during the COVID-19 pandemic [43]. Fisher et al memorably called LTCWs the "forgotten front line" [43]. Other researchers reported that LTCWs felt invisible and unsupported [44]. White et al noticed that although the media portrayed hospital staff as champions, negative media coverage of nursing homes was demoralizing [45]. We believe that our social web application and its co-development process demonstrate that LTCW perspectives truly matter and can shape the content, format, and moderation of a complex intervention designed to improve vaccine confidence among crucial health care workers.

## Next Steps

It is not yet certain that the COVID-19 Social Site will positively influence vaccine confidence or uptake. The results of our randomized trial will be available in 2024.

If successful, our intervention could become a template for other populations with low COVID-19 vaccine confidence or uptake. Additionally, this approach could suit different contexts where changing attitudes or behaviors could be beneficial for public health. Our web application is a light-touch, low-cost intervention that may be relatively easy to replicate and implement by governments, employers, and others. Additionally, the resources associated with launching web applications such as these could be scaled up or down. Without the constraints of a randomized trial, we could also replicate elements of this approach using established social media platforms.

## Conclusions

With user-centered design and participatory research, we developed a novel social media web application featuring curated internet content specifically for LTCWs.

## Acknowledgments

Research reported in this manuscript was funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (COVID-2021C2-13181). The statements presented in this manuscript are solely the responsibility of the authors and do not necessarily represent the views of the PCORI, or its Board of Governors or Methodology Committee. We acknowledge the broader CONFIDENT study team's support and advice, particularly our long-term care worker partners. We thank the following collaborators: Feng Chen, Patrick Delva, Jon Edwards, Branden Fillbrook, Tammy McIlnay, Katie Page, Lupita Rodriguez, Rowena Sheppard, Mary White and Celeste Wooten.

## Authors' Contributions

Conceptualization: CHS, GS, GE, MC, and MAD. Methodology: CHS, GS, GE, and MAD. Software: CM and AJ. Formal analysis: CHS, AS, JP, JH, and MAD. Resources: GE and MAD. Data curation: AS. Writing-original draft: CHS and AS. Writing-review and editing: all authors. Supervision: LJ, GS, GE, and MAD. Funding acquisition: GE and MAD.

## Conflicts of Interest

CHS holds copyright for the consideRATE suite of tools for serious illness. GE has edited and published books that provide royalties (Shared Decision Making, Oxford University Press and Groups, Radcliffe Press). GE's academic interests are focused on shared decision making and coproduction. He owns copyright in measures of shared decision making (collaboRATE) and care integration (integRATE), a measure of experience of care in serious illness (consideRATE), a measure of goal setting coopeRATE, a measure of clinician willingness to do shared decision making (incorpoRATE), and observer measures of shared decision making (Observer OPTION-5 and Observer OPTION-12). He is the Founder and Director of &think LLC, which owns the registered trademark for Option Grid patient decision aids. He is the Chief Clinical Research Scientist to abridge AI Inc. MAD has contributed to the development of Option Grid patient decision aids, and EBSCO Information Services sells subscription access to Option Grid patient decision aids. She receives consulting income from EBSCO Health and royalties. All other authors report no relevant conflicts.

Multimedia Appendix 1
Long-term care worker interview select quotes.
[DOCX File , 22 KB - jmir_v24i9e38359_app1.docx ]

Multimedia Appendix 2
Content identification methods and results.
[DOCX File , 490 KB - jmir_v24i9e38359_app2.docx ]

Multimedia Appendix 3
Long-term care worker interview methods and results.
[DOCX File , 21 KB - jmir_v24i9e38359_app3.docx ]

Multimedia Appendix 4
Long-term care worker interview guide.
[DOCX File , 17 KB - jmir_v24i9e38359_app4.docx ]

Multimedia Appendix 5
Long-term care worker content mix poll questionnaire.
[DOCX File , 304 KB - jmir_v24i9e38359_app5.docx ]

Multimedia Appendix 6
National online panel survey questionnaire.
[DOCX File , 16 KB - jmir_v24i9e38359_app6.docx ]

Multimedia Appendix 7
National online panel survey methods and results.

[DOCX File , 493 KB - jmir_v24i9e38359_app7.docx ]

---

Multimedia Appendix 8
Fact-checking methods.
[DOCX File , 20 KB - jmir_v24i9e38359_app8.docx ]

---

Multimedia Appendix 9
User experience testing methods and results.
[DOCX File , 21 KB - jmir_v24i9e38359_app9.docx ]

---

Multimedia Appendix 10
Community building methods and results.
[DOCX File , 38 KB - jmir_v24i9e38359_app10.docx ]

---

Multimedia Appendix 11
Abridged COVID-19 fact-checking list.
[DOCX File , 382 KB - jmir_v24i9e38359_app11.docx ]

---

## References

1. Guidance for the Interim Final Rule - Medicare and Medicaid Programs; Omnibus COVID-19 Health Care Staff Vaccination. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/files/document/qso-22-07-all.pdf [accessed 2022-08-06]
2. Nursing Home Staff Vaccination Rates Vary Widely by State as Vaccination Mandates Take Effect. Kaiser Family Foundation. 2022. URL: https://www.kff.org/coronavirus-covid-19/issue-brief/nursing-home-staff-vaccination-rates-vary-widely-by-state-as-vaccination-mandates-take-effect/ [accessed 2022-08-06]
3. Over 200,000 Residents and Staff in Long-Term Care Facilities Have Died From COVID-19. Kaiser Family Foundation. 2022. URL: https://www.kff.org/policy-watch/over-200000-residents-and-staff-in-long-term-care-facilities-have-died-from-covid-19/ [accessed 2022-08-06]
4. Van Houtven CH, DePasquale N, Coe NB. Essential long-term care workers commonly hold second jobs and double- or triple-duty caregiving roles. J Am Geriatr Soc 2020 Aug;68(8):1657-1660 [FREE Full text] [doi: 10.1111/jgs.16509] [Medline: 32338767]
5. Barry V, Dasgupta S, Weller DL, Kriss JL, Cadwell BL, Rose C, et al. Patterns in COVID-19 vaccination coverage, by social vulnerability and urbanicity - United States, December 14, 2020-May 1, 2021. MMWR Morb Mortal Wkly Rep 2021 Jun 04;70(22):818-824 [FREE Full text] [doi: 10.15585/mmwr.mm7022e1] [Medline: 34081685]
6. Nursing and Residential Care Facilities: NAICS 623. U.S. Bureau of Labor Statistics. URL: https://www.bls.gov/iag/tgs/iag623.htm [accessed 2022-08-06]
7. Survey: 94 Percent of Nursing Homes Face Staffing Shortages. American Health Care Association. URL: https://tinyurl.com/mp868277 [accessed 2022-08-06]
8. Horvath KJ, Ecklund AM, Hunt SL, Nelson TF, Toomey TL. Developing Internet-based health interventions: a guide for public health researchers and practitioners. J Med Internet Res 2015 Jan 23;17(1):e28 [FREE Full text] [doi: 10.2196/jmir.3770] [Medline: 25650702]
9. Daley MF, Narwaney KJ, Shoup JA, Wagner NM, Glanz JM. Addressing parents' vaccine concerns: A randomized trial of a social media intervention. Am J Prev Med 2018 Jul;55(1):44-54 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.010] [Medline: 29773490]
10. Glanz JM, Wagner NM, Narwaney KJ, Kraus CR, Shoup JA, Xu S, et al. Web-based social media intervention to increase vaccine acceptance: A randomized controlled trial. Pediatrics 2017 Dec;140(6):e20171117 [FREE Full text] [doi: 10.1542/peds.2017-1117] [Medline: 29109107]
11. Williams SE, Rothman RL, Offit PA, Schaffner W, Sullivan M, Edwards KM. A randomized trial to increase acceptance of childhood vaccines by vaccine-hesitant parents: a pilot study. Acad Pediatr 2013;13(5):475-480 [FREE Full text] [doi: 10.1016/j.acap.2013.03.011] [Medline: 24011750]
12. Ortiz RR, Smith A, Coyne-Beasley T. A systematic literature review to examine the potential for social media to impact HPV vaccine uptake and awareness, knowledge, and attitudes about HPV and HPV vaccination. Hum Vaccin Immunother 2019;15(7-8):1465-1475 [FREE Full text] [doi: 10.1080/21645515.2019.1581543] [Medline: 30779682]
13. Shoup JA, Wagner NM, Kraus CR, Narwaney KJ, Goddard KS, Glanz JM. Development of an interactive social media tool for parents with concerns about vaccines. Health Educ Behav 2015 Jun;42(3):302-312 [FREE Full text] [doi: 10.1177/1090198114557129] [Medline: 25413375]

14. Durand M, Alam S, Grande SW, Elwyn G. 'Much clearer with pictures': using community-based participatory research to design and test a Picture Option Grid for underserved patients with breast cancer. BMJ Open 2016 Feb 02;6(2):e010008 [FREE Full text] [doi: 10.1136/bmjopen-2015-010008] [Medline: 26839014]
15. Belsky V. Web application vs. website: finally answered. ScienceSoft. URL: https://www.scnsoft.com/blog/web-application-vs-website-finally-answered [accessed 2022-08-06]
16. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
17. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
18. Peretti-Watel P, Larson HJ, Ward JK, Schulz WS, Verger P. Vaccine hesitancy: clarifying a theoretical framework for an ambiguous notion. PLoS Curr 2015 Feb 25;7:ecurrents.outbreaks.6844c80ff9f5b273f34c91f71b7fc289 [FREE Full text] [doi: 10.1371/currents.outbreaks.6844c80ff9f5b273f34c91f71b7fc289] [Medline: 25789201]
19. Comparing Ways to Increase Confidence in COVID-19 Vaccines among Long-Term Care Workers -- The CONFIDENT Trial. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2021/confident-randomized-trial-increase-covid-19-vaccine-confidence-long-term-care-workers [accessed 2022-08-06]
20. Jarrett C, Wilson R, O'Leary M, Eckersberger E, Larson HJ, SAGE Working Group on Vaccine Hesitancy. Strategies for addressing vaccine hesitancy - A systematic review. Vaccine 2015 Aug 14;33(34):4180-4190 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.040] [Medline: 25896377]
21. COVID-19: Vaccine Options. DynaMed Decisions. URL: https://decisions.dynamed.com/shared-decision-making/covid-19-vaccine-options [accessed 2022-08-06]
22. Federal plain language guidelines. Plain Language. URL: https://plainlanguage.gov/guidelines/ [accessed 2019-03-27]
23. Braun V, Clarke V. Successful Qualitative Research A Practical Guide for Beginners. Thousand Oaks, CA, USA: SAGE Publications Ltd; 2013.
24. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. COVID-19 and Workers at Risk: Examining the Long-Term Care Workforce. Kaiser Family Foundation. 2020. URL: https://www.kff.org/report-section/covid-19-and-workers-at-risk-examining-the-long-term-care-workforce-tables/ [accessed 2022-08-06]
26. Fraud Detection. Qualtrics. URL: https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/ [accessed 2022-08-06]
27. Pre-Made Qualtrics Library Questions. Qualtrics. URL: https://www.qualtrics.com/support/survey-platform/survey-module/editing-questions/question-types-guide/pre-made-qualtrics-library-questions/
28. The American Community Survey. United States Census Bureau. 2021. URL: https://www2.census.gov/programs-surveys/acs/methodology/questionnaires/2021/quest21.pdf
29. Puckett JA, Brown NC, Dunn T, Mustanski B, Newcomb ME. Perspectives from Transgender and Gender Diverse People on How to Ask About Gender. LGBT Health 2020 Sep;7(6):305-311. [doi: 10.1089/lgbt.2019.0295]
30. ORARC Tip Sheet: Inclusive Demographic Data Collection. Office of Regulatory affairs and Research Compliance.: Harvard TH Chan School of Public Health URL: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/2102/2020/04/ORARC-Tip-Sheet-Inclusive-Demographic-Data-Collection.pdf
31. Jerman J, Jones RK, Onda T. Characteristics of U.S. Abortion Patients in 2014 and Changes Since 2008.: Guttmacher Institute; 2016. URL: https://www.guttmacher.org/report/characteristics-us-abortion-patients-2014
32. Donnelly KZ, Foster TC, Thompson R. What matters most? The content and concordance of patients' and providers' information priorities for contraceptive decision making. Contraception 2014 Sep 1;90(3):280-287. [doi: 10.1016/j.contraception.2014.04.012] [Medline: 24863169]
33. Why You Only Need to Test with 5 Users. Nielsen Norman Group. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2022-08-06]
34. Willis GB. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA, USA: SAGE Publications Ltd; 2005.
35. Allison R, Hayes C, McNulty CAM, Young V. A comprehensive framework to evaluate websites: Literature review and development of GoodWeb. JMIR Form Res 2019 Oct 24;3(4):e14372 [FREE Full text] [doi: 10.2196/14372] [Medline: 31651406]
36. Yale University. URL: https://usability.yale.edu/ [accessed 2022-08-06]
37. Writing an Effective Guide for a UX Interview. Nielsen Norman Group. URL: https://www.nngroup.com/articles/interview-guide/ [accessed 2022-08-06]
38. Lucero A. Using Affinity Diagrams to Evaluate Interactive Prototypes. In: Abascal J, Barbosa S, Fetter M, Gross T, Palanque P, Winckler M, editors. Human-Computer Interaction – INTERACT 2015. INTERACT 2015. Lecture Notes in Computer Science(), vol 9297. Cham: Springer; 2015:231-248. [doi: https://doi.org/10.1007/978-3-319-22668-2_19]

39. 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 Jul 09;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
40. Durand M, Yen RW, O'Malley AJ, Schubbe D, Politi MC, Saunders CH, et al. What matters most: Randomized controlled trial of breast cancer surgery conversation aids across socioeconomic strata. Cancer 2021 Feb 01;127(3):422-436 [FREE Full text] [doi: 10.1002/cncr.33248] [Medline: 33170506]
41. Crocker JC, Ricci-Cabello I, Parker A, Hirst JA, Chant A, Petit-Zeman S, et al. Impact of patient and public involvement on enrolment and retention in clinical trials: systematic review and meta-analysis. BMJ 2018 Nov 28;363:k4738 [FREE Full text] [doi: 10.1136/bmj.k4738] [Medline: 30487232]
42. Bird M, Ouellette C, Whitmore C, Li L, Nair K, McGillion MH, et al. Preparing for patient partnership: A scoping review of patient partner engagement and evaluation in research. Health Expect 2020 Jun;23(3):523-539 [FREE Full text] [doi: 10.1111/hex.13040] [Medline: 32157777]
43. Fisher E, Cárdenas L, Kieffer E, Larson E. Reflections from the "Forgotten Front Line": A qualitative study of factors affecting wellbeing among long-term care workers in New York City during the COVID-19 pandemic. Geriatr Nurs 2021;42(6):1408-1414 [FREE Full text] [doi: 10.1016/j.gerinurse.2021.09.002] [Medline: 34624698]
44. Sterling MR, Tseng E, Poon A, Cho J, Avgar AC, Kern LM, et al. Experiences of home health care workers in New York City during the coronavirus disease 2019 pandemic: A qualitative analysis. JAMA Intern Med 2020 Nov 01;180(11):1453-1459 [FREE Full text] [doi: 10.1001/jamainternmed.2020.3930] [Medline: 32749450]
45. White EM, Wetle TF, Reddy A, Baier RR. Front-line nursing home staff experiences during the COVID-19 pandemic. J Am Med Dir Assoc 2021 Jan;22(1):199-203 [FREE Full text] [doi: 10.1016/j.jamda.2020.11.022] [Medline: 33321076]

## Abbreviations

**LTCW:** long-term care worker
**UI:** user interface
**UX:** user experience

---

*Edited by M Gisondi; submitted 31.03.22; peer-reviewed by J Polo; comments to author 13.07.22; revised version received 19.07.22; accepted 03.08.22; published 22.09.22.*

*Please cite as:*
*Saunders CH, Sierpe A, Stevens G, Elwyn G, Cantrell M, Engel J, Gonzalez M, Hayward M, Huebner J, Johnson L, Jimenez A, Little NR, McKenna C, Onteeru M, Oo Khine M, Pogue J, Salinas Vargas JL, Schmidt P, Thomeer R, CONFIDENT Study Long-Term Care Partners, Durand MA*
*Co-Development of a Web Application (COVID-19 Social Site) for Long-Term Care Workers ("Something for Us"): User-Centered Design and Participatory Research Study*
*J Med Internet Res 2022;24(9):e38359*
*URL: https://www.jmir.org/2022/9/e38359*
*doi:10.2196/38359*
*PMID:35926074*

Original Paper

# Public Health Information Seeking, Trust, and COVID-19 Prevention Behaviors: Cross-sectional Study

Emmanuel Kwabena Tetteh[1], MPH, MD; Todd Combs[2], PhD; Elvin Hsing Geng[3], MPH, MD; Virginia Ruth McKay[2], PhD

[1]Office of Health Information and Data Science, Washington University School of Medicine in St Louis, St Louis, MO, United States
[2]Center for Public Health Systems Science, Washington University in St Louis, St Louis, MO, United States
[3]Division of Infectious Diseases, Washington University School of Medicine in St Louis, St Louis, MO, United States

**Corresponding Author:**
Virginia Ruth McKay, PhD
Center for Public Health Systems Science
Washington University in St Louis
One Brookings Drive
St Louis, MO, 63130
United States
Phone: 1 314 935 5740
Email: virginia.mckay@wustl.edu

## Abstract

**Background:** Preventative health measures such as shelter in place and mask wearing have been widely encouraged to curb the spread of the COVID-19 disease. People’s attitudes toward preventative behaviors may be dependent on their sources of information and trust in the information.

**Objective:** The aim of this study was to understand the relationship between trusting in COVID-19 information and preventative behaviors in a racially and politically diverse metropolitan area in the United States.

**Methods:** We conducted a web-based cross-sectional survey of residents in St. Louis City and County in Missouri. Individuals aged ≥18 years were eligible to participate. Participants were recruited using a convenience sampling approach through social media and email. The Health Belief Model and the Socioecological Model informed instrument development, as well as COVID-19–related questions from the Centers for Disease Control and Prevention. We performed an ordinary least squares linear regression model to estimate social distancing practices, perceptions, and trust in COVID-19 information sources.

**Results:** Of the 1650 eligible participants, the majority (n=1381, 83.7%) had sought or received COVID-19–related information from a public health agency, the Centers for Disease Control and Prevention, or both. Regression analysis showed a 1% increase in preventative behaviors for every 12% increase in trust in governmental health agencies. At their lowest levels of trust, women were 68% more likely to engage in preventative behaviors than men. Overall, those aged 18-45 years without vulnerable medical conditions were the least likely to engage in preventative behaviors.

**Conclusions:** Trust in COVID-19 information increases an individual’s likelihood of practicing preventative behaviors. Effective health communication strategies should be used to effectively disseminate health information during disease outbreaks.

***(J Med Internet Res 2022;24(9):e37846)*** doi:10.2196/37846

**KEYWORDS**

COVID-19; public health; health communication; trust and mistrust; disease prevention; health measure; health information; cross-sectional study; Health Belief Model

## Introduction

As the COVID-19 epidemic continues, preventative behaviors remain an important means to stemming the spread of the infection, despite the availability of effective vaccines and treatment modalities. This need is owed in part to the emergence of new variants of the COVID-19 virus and the low uptake of COVID-19 vaccines globally [1]. A growing body of evidence outlines the importance of nonpharmacological measures, such as restrictions on public gatherings, in controlling and preventing the spread of the disease [2-4]; however, it remains largely variable how willing people are to adopt these behaviors and for how long [5,6].

Initial research outlining the kinds of behaviors that individuals are the most likely to engage in to prevent the spread of COVID-19 is emerging. At the foundation are individuals' perceptions of risk for acquiring COVID-19 along with its expected severity for each person. Those who perceive themselves to be at a higher risk for acquiring COVID-19 and experiencing a poor outcome from the disease are more likely to engage in preventative behaviors [7-9]. Additionally, the source of information and the trustworthiness of the information is potentially critical. Earlier studies conducted in multiple settings have found that trust in COVID-19 health information from government officials and public health agencies (PHAs) was related to an increase in people's perceived level of risk, greater severity of the disease if infected, and greater belief in the effectiveness of preventive behaviors [10-17]. Political ideologies, religiosity, and conspiracy ideation have been identified to play a substantial role mediating trust in COVID-19 information and guidelines [13,15], highlighting the demographic differences influencing trust in health information and communication.

This study aimed to contribute to the growing body of evidence targeting the relationship between trust in COVID-19 information and preventative behaviors in St. Louis, Missouri, a relatively small metropolitan area with a racially and politically diverse population, where the burden of COVID-19 was slower to emerge than other major metropolitan areas. This study focused on the early window of the COVID-19 epidemic, prior to the availability of vaccines when state and local officials had imposed policies enacting a number of protective behaviors; however, the majority of the behaviors were voluntary, occurring even before policy measures started going into effect [4]. As of October 5, 2020, there were 32,589 confirmed COVID-19 cases in the St. Louis region including St. Louis City and County. Women had a higher rate of infection (2112 per 100,000) than men (1764 per 100,000). The disease had a 3.2% case fatality rate overall, but there were notable disparities by race—the rate was 2 times higher in Black or African American individuals (169.6 per 100,000) than White individuals (84.4 per 100,000) [18].

## Methods

### Study Design

We conducted a cross-sectional survey from April 23 to July 2, 2020, of St. Louis City and County residents to collect information about perceptions of the COVID-19 epidemic and social distancing behaviors. The survey was administered through Qualtrics and was available via an anonymous link. Participants were not given an incentive for participation; however, for every individual who participated in the survey, a US $1 donation was made to a local nonprofit organization working to counter the economic impacts of the epidemic in the St. Louis region, up to US $2,000. Data were collected from April 23 to July 2, 2020.

### Ethics Approval

The Institutional Review board of Washington University in St. Louis approved the study protocol and procedures of informed consent before the formal survey (#202004131). The Checklist for Reporting Results of Internet E-Surveys was used as a guide to report results and develop this manuscript [19].

### Participants

Individuals aged ≥18 years were eligible to participate in the study. Participants were recruited through targeted social media advertising and distribution through local email listserves. Participant recruitment continued throughout the data collection period. We aimed to collect a representative sample of St. Louis City and County residents based on the following variables: gender, age, socioeconomic status, and race/ethnicity. To increase the representativeness of our sample for St. Louis City and County residents, we constructed sample weights for the regression model. Using 5-year estimates from the American Community Survey (2015-2019) for the public use microdata areas encompassing the city and county [20], we used logistic regression weighting on samples of these data and of the survey data to calculate inverse probability weights. The sample was weighted on household income, race, gender, and age. Due to initially high variance in the weights, we trimmed them according to common practice to produce the final set [21,22].

### Measures

The survey asked participants a series of demographic questions (as shown below) about their zip code of residence, gender, age, ethnicity, race, employment status, social distancing policy, and comorbidities relevant to COVID-19. The survey included questions on comorbid conditions adapted from the Centers for Disease Control and Prevention (CDC) COVID-19 Community Survey Question Bank [23]. Perceptions of the COVID-19 epidemic and social distancing behaviors were measured using individual items corresponding to the major components of the Health Belief Model [24]. The Health Belief Model is a well-established framework that consists of 5 major components: likelihood of action, perceived threat, expected utility, self-efficacy, and cues to action. In the context of adopting preventative COVID-19 behaviors, an individual is likely to engage in social distancing if they perceive themselves to be at risk for COVID-19 (perceived threat), have adequate knowledge of social distancing (cues to action), feel that it will help reduce their risk (expected utility), and feel that they are able to participate in social distancing (self-efficacy) in the context of perceived benefits and barriers of action (expected utility). The survey was also informed by the Socioecological Model, which postulates that health behaviors are affected by factors that occur at individual, interpersonal, community, and societal levels [25]. Participants were asked to select the factors that influence their willingness and ability to engage in preventative behaviors. Lastly, participants were asked about perceptions of how COVID-19 information is communicated, adapted from the Health Information National Trends Survey 4 Cycle 1 instrument [26]. The Health Information National Trends Survey is a well-established, validated instrument that assesses the impact of the health information environment. We adapted questions A7 and A6, using a 4-point Likert scale, to measure participants' sources of COVID-19 information and trust in COVID-19 information sources, respectively. The full survey instrument is included as an appendix (Multimedia Appendix 1).

### Data Analyses

We downloaded data from Qualtrics and used R statistical software (version 4.0.1; R Foundation for Statistical Computing) for analysis. Descriptive statistical methods were used to summarize data on demographic characteristics. Categorical variables were summarized as frequencies (n) and percentages (%). For the main analyses, we performed an ordinary least squares linear regression model to estimate social distancing knowledge, perceptions, and practices. For the dependent variable, we constructed a preventive behaviors and attitudes (PBA) factor index of 12 reported practices and attitudes toward social distancing and other preventive behaviors, including hand washing, mask wearing, and knowledge and efficacy of social distancing behaviors (see the full list and distributions in Results). This index was operationalized as a proxy for the components of the Health Belief model. We also calculated a trust in public health institutions index from 2 items gauging trust in federal and state and local health agencies to serve as a predictor of social distancing practices and attitudes. The independent variables included this index, demographic characteristics (age, gender, income, race, employment status, and county of residence), and the presence or absence of preexisting health conditions that make individuals more vulnerable to COVID-19. Finally, we included individuals' perceptions of how likely they were to contract COVID-19 and 2 interaction terms: gender and the level of trust in PHAs; and age and the presence or absence of a preexisting condition that increases COVID-19 vulnerability. We hypothesized from previous literature that women would be more likely than men to take precautions if they had a high level of trust in PHAs [27,28] and that older adults with preexisting conditions would be more likely to take precautions than their younger counterparts [29]. All of these variables and their levels of measurement are described in the beginning of the Results section below.

## Results

### Participant Characteristics

The number of individuals responding to the survey between April 30 and July 2, 2020, was 3180. Among the respondents, 51.9% (n=1650) were aged ≥18 years and lived in St. Louis City or County and thus were eligible for analysis. Participant demographic characteristics are shown in Table 1.

Of the 1650 respondents, just over half (n=879, 53.3%) were aged 18-45 years, and 76.3% (n=1259) were women. Most (n=1426, 86.4%) respondents were White and 96 (5.8%) were Black or African American. More than half (n=912, 55.3%) reported annual household incomes of at least US $70,000, 22.7% (n=375) earned between US $40,000 and US $70,000, 13.6% (n=225) reported incomes less than US $40,000, and 8.4% (n=138) did not respond. About half (805/1650, 48.8%) reported currently working from home, 32.1% (529/1650) were not working, and 18.3% (302/1650) currently worked outside the home.

**Table 1.** Demographics and characteristics.

| Characteristic | Respondent (N=1650), n (%) |
|---|---|
| **Age (years)** | |
| 18-25 | 98 (5.9) |
| 26-35 | 368 (22.3) |
| 36-45 | 413 (25) |
| 46-55 | 259 (15.7) |
| 56-65 | 284 (17.2) |
| ≥66 | 227 (13.8) |
| **Gender** | |
| Gender nonconforming | 29 (1.8) |
| Man | 346 (21) |
| Woman | 1259 (76.3) |
| No response | 16 (1) |
| **Race/ethnicity** | |
| Asian | 29 (1.8) |
| Black or African American | 96 (5.8) |
| Hispanic or Latino | 27 (1.6) |
| Multiple races or ethnicities | 53 (3.2) |
| Other | 16 (1) |
| White | 1426 (86.4) |
| No response | 3 (0.2) |
| **Household income (US $)** | |
| <20,000 | 62 (3.8) |
| 20,000 to <30,000 | 78 (4.7) |
| 30,000 to <40,000 | 85 (5.2) |
| 40,000 to <50,000 | 130 (7.9) |
| 50,000 to <70,000 | 245 (14.8) |
| 70,000 to <100,000 | 324 (19.6) |
| 100,000 to <150,000 | 312 (18.9) |
| ≥150,000 | 276 (16.7) |
| No response | 138 (8.4) |
| **Employment status** | |
| Working from home | 805 (48.8) |
| Not working | 529 (32.1) |
| Working outside the home | 302 (18.3) |
| No response | 14 (0.8) |
| **COVID-19–vulnerable health conditions**[a] | |
| At least 1 | 534 (32.3) |
| None | 1116 (67.6) |

[a] Respondents were asked about asthma; cancer; chronic heart, kidney, and lung diseases; diabetes; and immunosuppressive conditions.

### Sources of COVID-19 Information

Of the 1650 respondents, most (n=1381, 83.7%) sought out or received information about COVID-19 from PHAs, including local, state, and national PHAs (Table 2). In all, 58.7% (n=969) of respondents had sought or received information from both a local or state health department and the CDC, 3.4% (n=56) had done so only from state or local PHAs, 21.6% (n=356) had gotten information solely from the CDC, and 16.3% (n=269) had not received any information from a PHA. Most respondents had a moderate (n=751, 45.5%) or high (n=512, 31%) amount of trust in federal PHAs, and the remaining 23.2% (n=382) had little or no trust in federal PHAs (Table 2). Similarly, for local or state PHAs, most had a moderate (n=801, 48.5%) or high (n=495, 30%) amount of trust, and 21% (n=347) had little or no trust.

**Table 2.** Sources of information, trust in public health agencies, and perceptions of risk.

| Topic | Respondent (N=1650), n (%) |
|---|---|
| **Sources of information** | |
| None | 269 (16.3) |
| State or local PHA[a] | 56 (3.4) |
| CDC[b] | 356 (21.6) |
| State or local PHA and CDC | 969 (58.7) |
| **Trust in federal PHAs** | |
| Not at all | 100 (6.1) |
| A little | 282 (17.1) |
| A moderate amount | 751 (45.5) |
| A lot | 512 (31) |
| No response | 5 (0.3) |
| **Trust in state or local PHAs** | |
| Not at all | 61 (3.7) |
| A little | 286 (17.3) |
| A moderate amount | 801 (48.5) |
| A lot | 495 (30) |
| No response | 7 (0.4) |
| **Likelihood of contracting COVID-19** | |
| Unlikely | 626 (37.9) |
| Neither | 498 (30.2) |
| Likely | 517 (31.3) |

[a] PHA: public health agency.

[b] CDC: Centers for Disease Control and Prevention.

### Perceptions of Risk and Testing

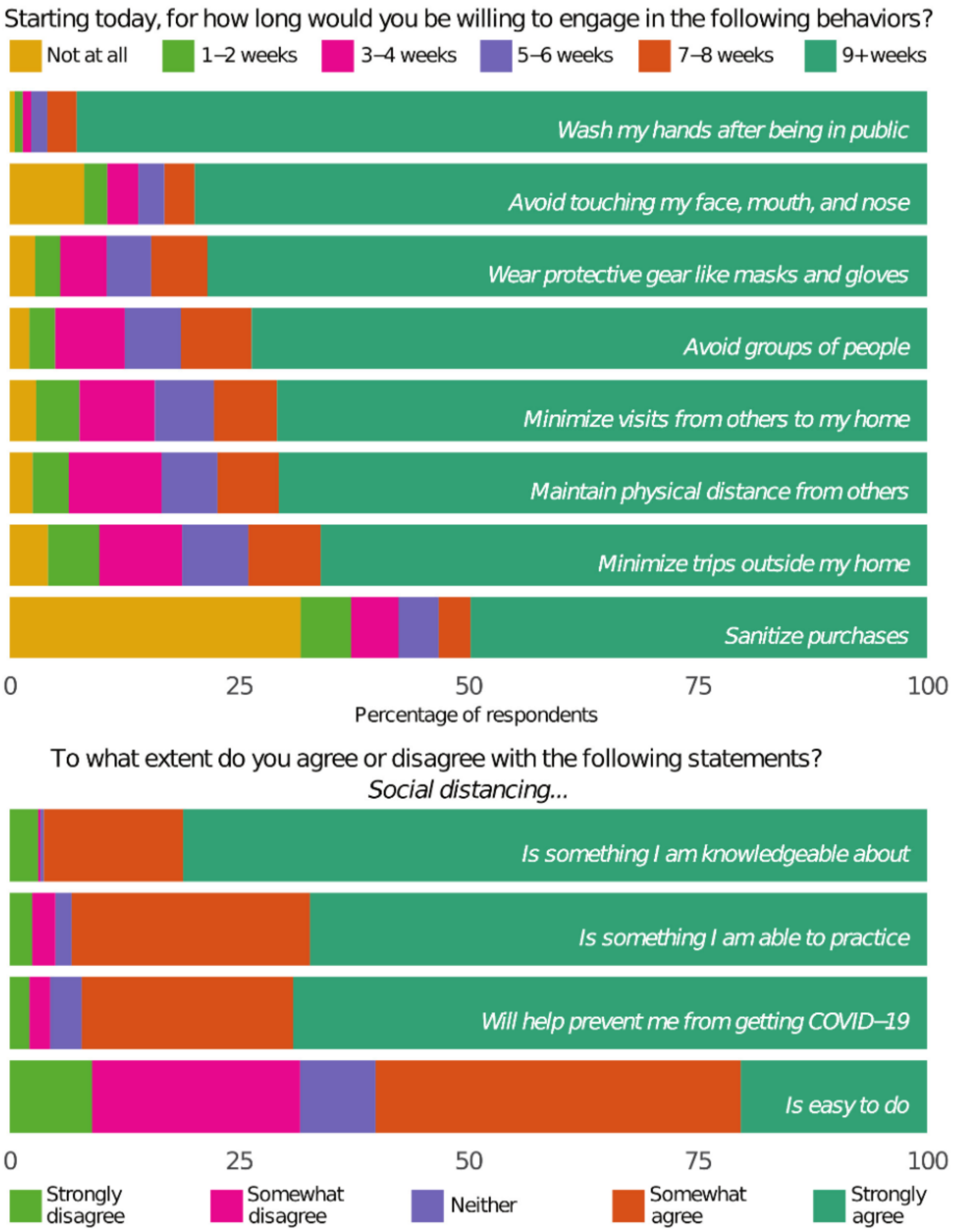
Table 2 also shows the perceptions of risk in the population. Of the 1650 respondents, about one-third (n=517, 31.3%) of participants thought that they were likely to contract COVID-19 in the next 3 months, and a larger percentage (n=626, 37.9%) responded that they were unlikely to contract COVID-19 in the next 3 months. The remaining 30.2% (n=498) responded that they were neither likely nor unlikely. About one-third (n=534, 32.3%) of respondents reported preexisting health conditions that made them more likely to contract or experience moderate or severe cases of COVID-19 (eg, asthma and chronic heart, kidney, or lung disease).

### Preventive Behaviors and Social Distancing

The survey also asked about which social distancing and other preventive behaviors the respondents were engaging in and how long they were willing to do so (Figure 1). Of the 1650 respondents, the clear majority—at least 75%—were willing to engage in 4 of the 8 specific behaviors for 9 weeks or more: washing hands after being in public (n=1512, 91.6%), avoiding touching one’s face (n=1298, 78.7%), wearing protective gear (n=1282, 77.7%), and avoiding groups (n=1203, 72.9%). Between half and three-quarters of the respondents were willing to engage in the other 4 behaviors for 9 weeks or more: minimizing in-home visitors (n=1155, 70.9%) and trips from home (n=1080, 65.5%), maintaining physical distance from others (n=1161, 70.4%), and sanitizing purchased goods (n=797, 46.5%). The majority of respondents either strongly (n=1331, 80.7%) or somewhat (n=249, 15.1%) agreed with the statement

"I am knowledgeable about social distancing." Similarly, 66.5% (n=1098) strongly agreed and 25.7% (n=424) somewhat agreed that they were able to practice social distancing, and a comparable number of respondents agreed somewhat (n=377, 22.8%) or strongly (n=1129, 68.4%) that social distancing would help prevent COVID-19 transmission. However, when asked whether social distancing was easy to do, only 20.2% (n=333) strongly and 39.6% (n=653) somewhat agreed, with 22.5% (n=372) somewhat and 8.9% (n=147) strongly disagreeing with the statement.

**Figure 1.** Willingness to practice and attitudes toward preventative behaviors.



## Linear Regression

To model whether respondents' actions and attitudes were influenced by demographics, sources of COVID-19 information, trust in those sources, and perceptions of risk for getting COVID-19, we constructed a factor index of the 12 survey items in Figure 1. The PBA index had a Cronbach α of .83 (95% CI .81-.84) and ranged from 0.41 to 5. The mean value was 4.2 (SD 0.82), and the natural log of this index was used as the outcome variable to approximate linearity. A higher score on the index means more practice of preventative behaviors. We also constructed a trust in public institutions index from the 2 survey items on trust in information from local or state and federal PHAs to serve as a predictor variable. The trust index had a Cronbach α of .80 (95% CI .79-.82) and ranged from 0 to 3, with a mean of 2.0 (SD 0.75). A higher score on the trust index means more trust in a PHA. From April 30 to July 2, 2020, both of these indexes stayed relatively constant (Figure 2), with daily reported averages around the overall mean for each.

Table 3 presents the results of the linear regression model. As explained above in the Methods section, the model used inverse probability weights to increase the representativeness of the populations of St. Louis City and County. Although the variance of the initial weights was relatively high (0.4; range 3.0-5.5), after trimming, it was 0.27 (range 3.0-5.2). The outcome of the model—the (natural log of the) PBA index—was regressed on 3 main categories of variables: demographics, trust in public health institutions, and individual risk perceptions. In addition, 2 interaction terms were included as moderators, and a control variable for time gauged the evolution of preventive behaviors throughout the 10 weeks of data collection. We executed the model in R statistical software (version 4.0.1) using the *svyglm* function in the *survey* package [30], which calculates robust standard errors to account for the weights included in the model.

**Figure 2.** Average preventative behaviors and public health trust indices throughout data collection: April 30 to July 2.

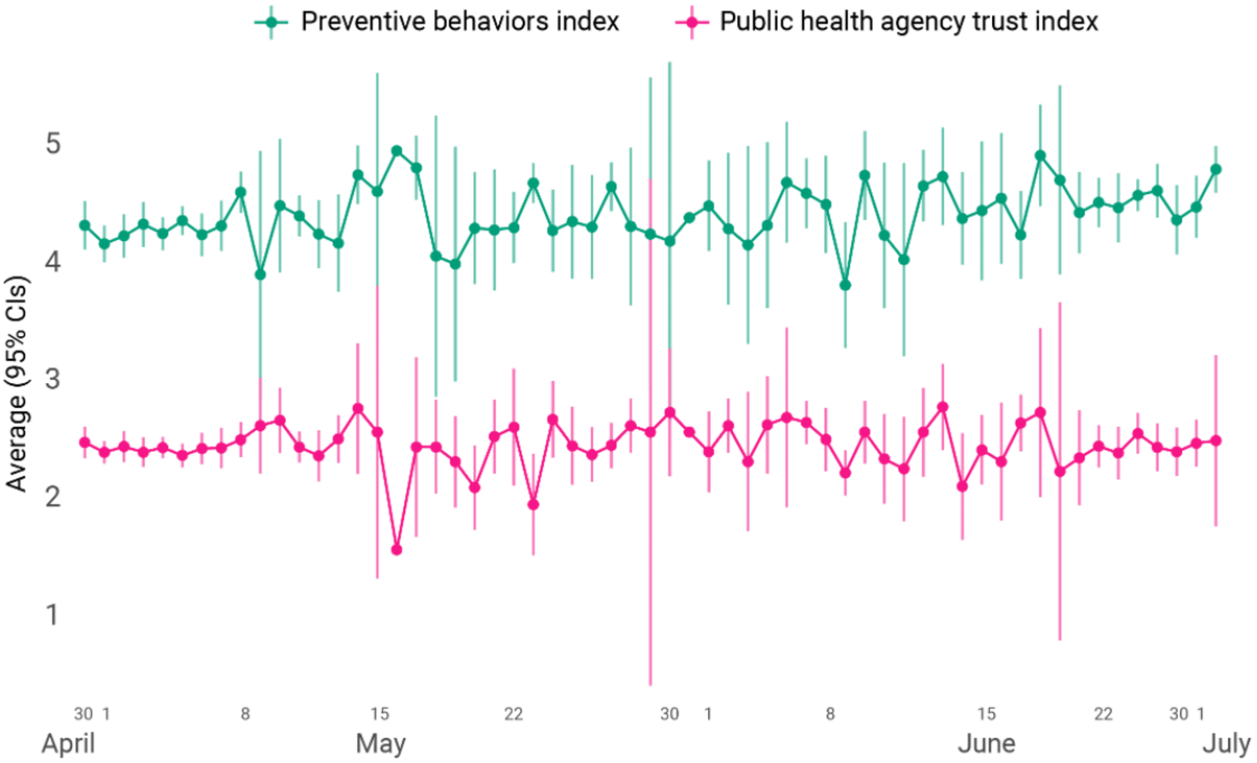


**Table 3.** Linear regression results: the effect of demographic characteristics, risk perception, and public health agency trust on COVID-19 preventative attitudes and behaviors (N=1440; adjusted $R^2$=0.12). Outcome is the natural log of the factor index of willingness to and attitudes toward preventive behaviors. CIs were calculated with heteroskedasticity-robust standard errors. When checked for multicollinearity, the maximum variance inflation factor value was 1.09 for the age predictor.

| Variable | OLS[a] coefficient | 95% CI | *P* value |
|---|---|---|---|
| **Demographics** | | | |
| **Age (years)** | | | |
| 18-45 | Reference | Reference | |
| 46-65 | 0.03 | 0.00 to 0.06 | .05 |
| ≥66 | 0.05 | 0.00 to 0.09 | .10 |
| **Gender** | | | |
| Women | Reference | Reference | |
| Men | –0.68 | –1.06 to –0.29 | <.001 |
| **Income (US $)** | | | |
| <40,000 | Reference | Reference | |
| 40,000 to <70,000 | 0.04 | 0.00 to 0.08 | .05 |
| ≥70,000 | 0.02 | –0.00 to 0.06 | .16 |
| **Race/ethnicity** | | | |
| Black or African American | –0.01 | –0.05 to 0.04 | .85 |
| Other races/ethnicities | 0 | –0.05 to 0.06 | >.99 |
| White | Reference | Reference | |
| **Employment status** | | | |
| Working outside the home | Reference | Reference | |
| Working from home | 0.05 | 0.02 to 0.08 | <.001 |
| Not working | 0.05 | 0.01 to 0.08 | .01 |
| **County of residence** | | | |
| St. Louis City | Reference | Reference | |
| St. Louis County | –0.02 | –0.05 to –0.00 | .04 |
| **Vulnerable conditions** | | | |
| None | Reference | Reference | |
| At least 1 | 0.05 | 0.02 to 0.08 | .01 |
| **Public health institutions** | | | |
| Trust in PHAs[b] (index, log-transformed) | 0.12 | 0.02 to 0.22 | .03 |
| Trust in doctors | 0.03 | –0.01 to 0.06 | <.001 |
| Amount of evidence from PHAs | 0.01 | 0.00 to 0.03 | .10 |
| **Likelihood of getting COVID-19** | | | |
| **Individual perception** | | | |
| Not likely | Reference | Reference | |
| Neither | 0.01 | –0.03 to 0.03 | .96 |
| Likely | 0.02 | –0.01 to 0.05 | .11 |
| **Interaction terms** | | | |
| **Gender and trust in PHAs** | | | |
| Women | Reference | Reference | |
| Men | 0.52 | 0.22 to 0.82 | <.001 |

| Variable | OLS[a] coefficient | 95% CI | *P* value |
|---|---|---|---|
| **Age (years) and vulnerable conditions** | | | |
| 18-45 | Reference | Reference | |
| 46-65 | –0.06 | –0.11 to 0.00 | .05 |
| ≥66 | –0.05 | –0.12 to 0.02 | .24 |
| **Time** | | | |
| Week number (April 30 to July 2) | 0.02 | 0.00 to 0.01 | <.001 |

[a] OLS: ordinary least squares.

[b] PHA: public health agency.

Since the dependent variable was log-transformed to better approximate linearity, the interpretation of the coefficients requires an extra step. For all but trust in PHAs index, which was also log-transformed, the formula for back-transforming the coefficient estimates is as follows:

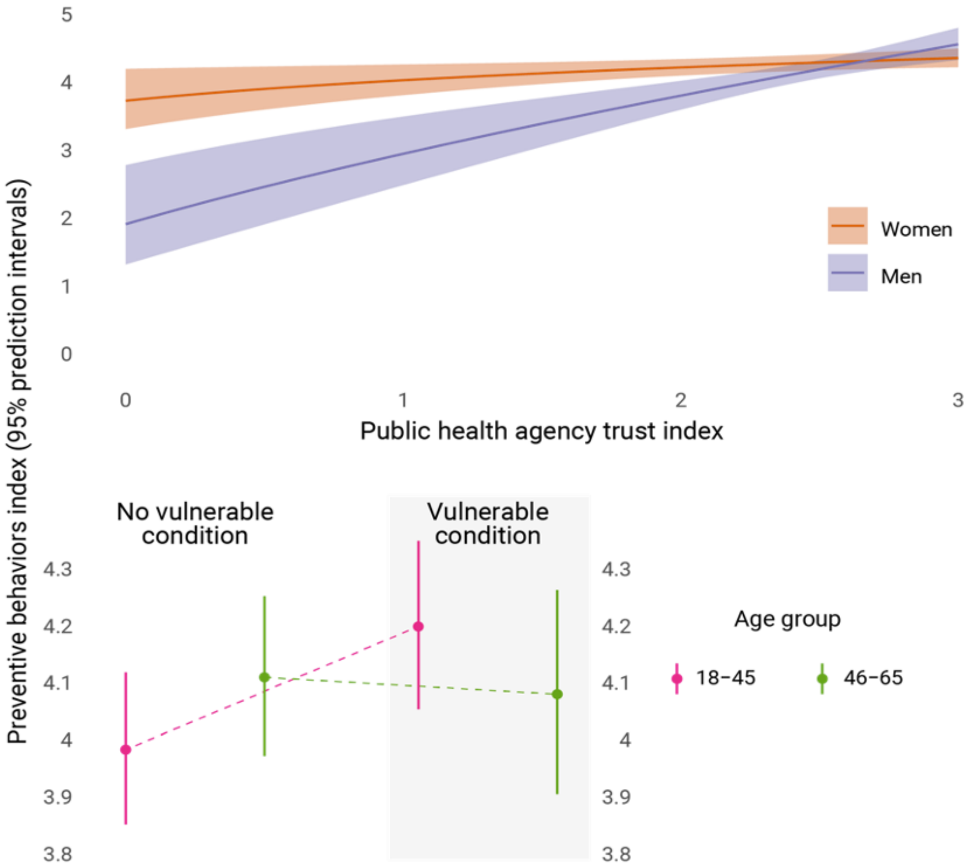
where $\beta_x$ is the estimated coefficient. Once transformed using this formula, the coefficients are approximately multiplied by 100.

For demographic characteristics, one difference was found between the 3 age groups. Those aged 46-65 years scored 3% higher on the PBA index than their younger (aged 18-45 years) counterparts (*P*=.05), all else equal. Women in our sample were estimated to have PBA index scores two-thirds (68%) higher than men on average (*P*<.001), and scores for middle-income individuals reporting between US $40,000 and US $70,000 in annual household income were 4% higher than those reporting less than US $40,000 (*P*=.05). No statistically significant differences were found between individuals who were White, Black or African American (*P*=.85), and of other races or ethnicities (*P*>.99). Respondents who were not working and those who were working from home both scored 5% higher than those working outside the home for preventive behaviors (*P*<.001 and *P*=.03, respectively). Residents of St. Louis County scored 2% lower than those in the city on average (*P*=.04), and those with at least 1 COVID-19–vulnerable health condition reported 5% higher social distancing attitudes and practices than those with no related conditions (*P*=.01).

The relationship between trust in PHAs and preventive behaviors was elastic, since both the independent and dependent variables were log-transformed, and can be interpreted as for every 12% increase in trust, a 1% increase in precautionary practices resulted. Trust in PHAs also moderated the relationship between respondents' gender and preventive behaviors and attitudes, as increased trust among men narrowed the gap between the genders by 0.52% as further illustrated below (*P*<.001). To a lesser extent, the presence of vulnerable conditions led to decreased differences (–6%) between respondents aged 46-65 and 18-45 years in preventive behaviors (*P*=.05). Trust in doctors, the amount of COVID-19 evidence received from PHAs, and individual risk perceptions were not related to preventive behaviors after controlling for the effects of all other variables in the model. Finally, the positive and statistically significant effect of the variable for week completing the survey suggested a slight 2% increase on average over time (*P*<.001).

To further illustrate how trust in PHAs led to decreased differences across gender, the top panel in Figure 3 shows the average predictions for women and men along the range of trust. Although the largest differences in preventive behaviors and attitudes is present at the lowest levels of trust—women at about 4 on the index versus men at 2.7, representing a 67% difference—as trust in PHAs increases among men, the differences narrow and eventually disappear at the highest trust levels. The bottom panel of Figure 3 considers the impact of preexisting COVID-19–vulnerable conditions on the differences between those aged 18-45 and 46-65 years for preventive behaviors. For those with no such conditions, the younger group is about 6% lower on the index. Those aged 18-45 years with vulnerable conditions have preventive behaviors and attitudes 5% higher than the same group without conditions and are 3% higher on the index than their older counterparts who also have vulnerable conditions. The relative scores for preventive behaviors flip between the 2 age groups when comparing those with and without vulnerable conditions.

**Figure 3.** Model estimates for practicing behaviors (scale: 0=low to 5=high) for interaction terms (top: gender and public health trust; bottom: age and vulnerable health conditions). Results were calculated using the average values of all other covariates.



## Discussion

### Principal Findings

The purpose of this study was to investigate the relationship between trust in COVID-19 information and engaging in preventive behaviors among residents in the St. Louis region, including St. Louis City and County, in the early window of the epidemic prior to the development of vaccines and treatment modalities. The majority of respondents had sought or received COVID-19–related information from a PHA and trusted that information. Those who expressed trust in the information from PHAs were more likely to engage in preventative behaviors. Our results show that PHAs are still an important source of information in disease outbreaks, and contrary to the vocalization of people not obeying [31,32], the majority of people still listen to their PHAs. Across all demographic groups, preventative behaviors improved as trust increased. In our sample, people's trust in sources of information and their practice of preventative behaviors remained relatively consistent with a slight 2% increase on average throughout the period of data collection, regardless of the changes in the severity of the disease (ie, caseload and case fatality rate) in the region.

This study contributes to the limited scientific literature regarding the association between COVID-19 preventative behaviors, the trustworthiness of information, and sources of information. Our findings are comparable to earlier studies that found that people who had higher trust in government COVID-19 messaging were more likely to adopt preventive behaviors [7,10-12,14,15]. In alignment with existing work, our respondents felt that preventive measures, such as social distancing, would help prevent the disease spread, but only a minority of people strongly agreed that social distancing was easy to do [33]. Public health campaigns are usually implemented under the assumption that once information is disseminated and knowledge is enhanced, recommended behaviors will follow [34]. Our findings demonstrate that this assumption is not always valid; rather, people's perception of risk and their ability to engage in preventive behaviors (self-efficacy) are more likely to influence their health behavior [35,36].

A potential threat to people's accurate perception of COVID-19 risk is through misinformation. Although it is unclear to what extent misinformation among our study population impacted our results, other studies have reported the prevalence of widespread false information about the COVID-19 disease and the effect of misinformation on people's perceived risk and adoption of preventative behaviors [37,38]. Results from a study conducted in 52 countries showed that 83% of vaccine-related rumors on popular web-based platforms were false [38], posing a substantial threat to vaccine uptake. It is incumbent upon PHAs, clinicians, and health practitioners to ensure that the most accurate and up-to-date disease risk information and preventative measures are carefully distilled and communicated to the public. Furthermore, it is imperative for public health messaging to debunk misleading and false information about the disease, modes of transmission, and the effectiveness of treatment and preventative measures.

Additionally, our study highlights a need for audience-targeted health communication that can effectively encourage different groups of people within a given population—specifically young people, men, and lower-income populations—to increase trust in the health information provided. Segmenting audience according to various demographic characteristics and behavioral traits increases the effectiveness of health communication campaigns intended to promote health behaviors [39]. The Risk Perception Attitude framework—a tool for assessing individuals' health behaviors based on their perceived risk and efficacy—can be used to guide health communication to different groups of people [40].

## Limitations

This study used a cross-sectional study design, which potentially limits the generalizability and representativeness of the results. The study population was not representative of the actual population, with some demographic groups substantially underrepresented, therefore reducing the generalizability of the results. This limitation was addressed by constructing sample weights for the regression model. The sample was weighted on household income, race, gender, and age, and our large sample size helped ensure statistical power. Since this study recruited participants voluntarily, people who were concerned by the COVID-19 disease or had been affected by it may have been more likely to participate, thus introducing a possible selection bias. Lastly, the survey responses were self-reported and may have led to some recall bias.

## Conclusion

This study provided insights into how preventive behaviors during the COVID-19 epidemic are influenced by sources of health-related information and the trustworthiness of information. We found that contrary to the vocalization of people not obeying, the majority of people still listen to their PHAs. Receiving health information from PHAs—and trusting that information—increased an individual's likelihood of engaging in preventative health behaviors. Incoherent COVID-19 information from state and local PHAs and blanket approaches to communicating health information have a decreased impact in addressing risk perceptions and efficacy beliefs in specific subpopulations, such as among men and young adults. Future research should consider how audience-targeted health communication strategies can ensure that different subpopulations adopt preventive health behaviors in disease outbreaks. Furthermore, PHAs and clinicians should make a continuous effort to debunk false and misleading COVID-19 information that may be prevalent on the internet and social media.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
The full COVID-19 survey instrument.
[PDF File (Adobe PDF File), 223 KB - jmir_v24i9e37846_app1.pdf ]

## References

1. Haque A, Pant AB. Mitigating COVID-19 in the face of emerging virus variants, breakthrough infections and vaccine hesitancy. J Autoimmun 2022 Feb;127:102792 [FREE Full text] [doi: 10.1016/j.jaut.2021.102792] [Medline: 34995958]
2. Askitas N, Tatsiramos K, Verheyden B. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. Sci Rep 2021 Jan 21;11(1):1972 [FREE Full text] [doi: 10.1038/s41598-021-81442-x] [Medline: 33479325]
3. Bo Y, Guo C, Lin C, Zeng Y, Li HB, Zhang Y, et al. Effectiveness of non-pharmaceutical interventions on COVID-19 transmission in 190 countries from 23 January to 13 April 2020. Int J Infect Dis 2021 Jan;102:247-253 [FREE Full text] [doi: 10.1016/j.ijid.2020.10.066] [Medline: 33129965]
4. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong social distancing measures in the United States reduced the COVID-19 growth rate. Health Aff (Millwood) 2020 Jul 01;39(7):1237-1246. [doi: 10.1377/hlthaff.2020.00608] [Medline: 32407171]
5. Moore RC, Lee AY, Hancock JT, Halley MC, Linos E. Age-related differences in experiences with social distancing at the onset of the COVID-19 pandemic: a computational and content analytic investigation of natural language from a social media survey. JMIR Hum Factors 2021 Jun 09;8(2):e26043 [FREE Full text] [doi: 10.2196/26043] [Medline: 33914689]
6. Resnick B, Brian. Social distancing can't last forever. here's what should come next. Vox. 2020 Mar 30. URL: https://www.vox.com/science-and-health/2020/3/26/21192211/coronavirus-covid-19-social-distancing-end [accessed 2022-02-28]
7. Azlan AA, Hamzah MR, Sern TJ, Ayub SH, Mohamad E. Public knowledge, attitudes and practices towards COVID-19: a cross-sectional study in Malaysia. PLoS One 2020 May 21;15(5):e0233668 [FREE Full text] [doi: 10.1371/journal.pone.0233668] [Medline: 32437434]
8. Clements JM. Knowledge and behaviors toward COVID-19 among US residents during the early days of the pandemic: cross-sectional online questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]

9. Motta Zanin G, Gentile E, Parisi A, Spasiano D. A preliminary evaluation of the public risk perception related to the COVID-19 health emergency in Italy. Int J Environ Res Public Health 2020 Apr 27;17(9):3024 [FREE Full text] [doi: 10.3390/ijerph17093024] [Medline: 32349253]
10. Ali SH, Foreman J, Tozan Y, Capasso A, Jones AM, DiClemente RJ. Trends and predictors of COVID-19 information sources and their relationship with knowledge and beliefs related to the pandemic: nationwide cross-sectional study. JMIR Public Health Surveill 2020 Oct 08;6(4):e21071 [FREE Full text] [doi: 10.2196/21071] [Medline: 32936775]
11. Han Q, Zheng B, Cristea M, Agostini M, Bélanger JJ, Gützkow B, PsyCorona Collaboration, et al. Trust in government regarding COVID-19 and its associations with preventive health behaviour and prosocial behaviour during the pandemic: a cross-sectional and longitudinal study. Psychol Med 2021 Mar 26:1-11 [FREE Full text] [doi: 10.1017/S0033291721001306] [Medline: 33769242]
12. Ipsen C, Myers A, Sage R. A cross-sectional analysis of trust of information and COVID-19 preventative practices among people with disabilities. Disabil Health J 2021 Apr;14(2):101062 [FREE Full text] [doi: 10.1016/j.dhjo.2021.101062] [Medline: 33495098]
13. Lermer E, Hudecek MFC, Gaube S, Raue M, Batz F. Early and later perceptions and reactions to the COVID-19 pandemic in Germany: on predictors of behavioral responses and guideline adherence during the restrictions. Front Psychol 2021 Nov 26;12:769206 [FREE Full text] [doi: 10.3389/fpsyg.2021.769206] [Medline: 34899517]
14. Neureiter A, Stubenvoll M, Kaskeleviciute R, Matthes J. Trust in science, perceived media exaggeration about COVID-19, and social distancing behavior. Front Public Health 2021 Dec 1;9:670485 [FREE Full text] [doi: 10.3389/fpubh.2021.670485] [Medline: 34926361]
15. Plohl N, Musil B. Modeling compliance with COVID-19 prevention guidelines: the critical role of trust in science. Psychol Health Med 2021 Jan;26(1):1-12. [doi: 10.1080/13548506.2020.1772988] [Medline: 32479113]
16. Schumpe BM, van Lissa CJ, Bélanger JJ, Ruggeri K, Mierau J, Nisa CF, et al. Predictors of adherence to public health behaviors for fighting COVID-19 derived from longitudinal data. Sci Rep 2022 Mar 09;12(1):3824 [FREE Full text] [doi: 10.1038/s41598-021-04703-9] [Medline: 35264597]
17. Seale H, Heywood AE, Leask J, Sheel M, Thomas S, Durrheim DN, et al. COVID-19 is rapidly changing: examining public perceptions and behaviors in response to this evolving pandemic. PLoS One 2020 Jun 23;15(6):e0235112 [FREE Full text] [doi: 10.1371/journal.pone.0235112] [Medline: 32574184]
18. Public Health Dashboard. Missouri Department of Health & Senior Services. 2020. URL: https://showmestrong.mo.gov/data/public-health/ [accessed 2020-11-11]
19. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Public Use Microdata Sample (PUMS). United States Census Bureau. 2021. URL: https://www.census.gov/programs-surveys/acs/microdata.html [accessed 2021-03-08]
21. Kalton G, Flores-Cervantes I. Weighting methods. J Off Stat 2003;19(2):81-97 [FREE Full text]
22. Kasprzyk D, Lepkowski JM, Kalton G. Weighting adjustments for partial nonresponse in the 1984 SIPP panel. Institute for Social and Economic Research (ISER). URL: https://www.iser.essex.ac.uk/research/publications/497269 [accessed 2021-04-01]
23. Form: CDC COVID-19 Community Survey Question Bank (DRAFT). Centers for Disease Control and Prevention. URL: https://cde.nlm.nih.gov/formView?tinyId=Kcceysolt [accessed 2020-11-11]
24. Kirscht JP. The Health Belief Model and illness behavior. Health Educ Monogr 1974 Dec 01;2(4):387-408. [doi: 10.1177/109019817400200406]
25. Bronfenbrenner U. Ecological systems theory. In: Vasta R, editor. Annals of Child Development: Vol 6. London, UK: Jessica Kingsley Publishers; 1989:187-249.
26. HINTS, Cycle 4, Full Content English Instrument, Annotated. National Cancer Institute. 2020. URL: https://hints.cancer.gov/docs/Instruments/HINTS5_Cycle4_AnnotatedInstrumentEnglish.pdf [accessed 2022-09-09]
27. Pavela Banai I, Banai B, Mikloušić I. Beliefs in COVID-19 conspiracy theories, compliance with the preventive measures, and trust in government medical officials. Curr Psychol 2021 May 26:1-11 [FREE Full text] [doi: 10.1007/s12144-021-01898-y] [Medline: 34075284]
28. O'Hara L, Abdul RH, Shi Z. Gender and trust in government modify: the association between mental health and stringency of public health measures to reduce COVID-19. 2020 Presented at: Qatar University Annual Research an Exhibition QUARFE 2020; October 27, 2020; Qatar. [doi: 10.29117/quarfe.2020.0282]
29. Ruiz JB, Bell RA. Predictors of intention to vaccinate against COVID-19: results of a nationwide survey. Vaccine 2021 Feb 12;39(7):1080-1086 [FREE Full text] [doi: 10.1016/j.vaccine.2021.01.010] [Medline: 33461833]
30. Lumley T. Analysis of complex survey samples. J Stat Soft 2004 Apr 15;9(8):1-19. [doi: 10.18637/jss.v009.i08]
31. Pollard MS, Davis LM. Decline in trust in the Centers for Disease Control and Prevention during the COVID-19 pandemic. Rand Health Q 2022 Jun;9(3):23 [FREE Full text] [Medline: 35837520]
32. Robert Wood Johnson Foundation. The Public's perspective on the United States public health system. Harvard T.H. Chan School of Public Health. 2021 May. URL: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/94/2021/05/RWJF-Harvard-Report_FINAL-051321.pdf [accessed 2022-02-07]

33. Betsch C. How behavioural science data helps mitigate the COVID-19 crisis. Nat Hum Behav 2020 May 27;4(5):438-438 [FREE Full text] [doi: 10.1038/s41562-020-0866-1] [Medline: 32221514]
34. Rimal RN. Perceived risk and efficacy beliefs as motivators of change: use of the Risk Perception Attitude (RPA) framework to understand health behaviors. Hum Commun Res 2003 Jul 01;29(3):370-399. [doi: 10.1093/hcr/29.3.370]
35. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
36. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
37. Kim HK, Tandoc EC. Consequences of online misinformation on COVID-19: two potential pathways and disparity by eHealth literacy. Front Psychol 2022 Feb 14;13:783909 [FREE Full text] [doi: 10.3389/fpsyg.2022.783909] [Medline: 35237207]
38. Islam MS, Kamal AM, Kabir A, Southern DL, Khan SH, Hasan SMM, et al. COVID-19 vaccine rumors and conspiracy theories: the need for cognitive inoculation against misinformation to improve vaccine adherence. PLoS One 2021 May 12;16(5):e0251605 [FREE Full text] [doi: 10.1371/journal.pone.0251605] [Medline: 33979412]
39. Williams JE, Flora JA. Health behavior segmentation and campaign planning to reduce cardiovascular disease risk among Hispanics. Health Educ Q 1995 Feb 04;22(1):36-48. [doi: 10.1177/109019819502200105] [Medline: 7721600]
40. Thompson TL, editor. Risk Perception Attitude Framework. In: Encyclopedia of Health Communication. New York, NY: SAGE Publications; Sep 10, 2014.

## Abbreviations

**CDC:** Centers for Disease Control and Prevention
**PBA:** preventive behaviors and attitudes
**PHA:** public health agency

*Edited by M Gisondi, M Gottlieb; submitted 09.03.22; peer-reviewed by S Governor, JS Tham, R Meng; comments to author 19.04.22; revised version received 28.07.22; accepted 03.08.22; published 30.09.22.*

*Please cite as:*
*Tetteh EK, Combs T, Geng EH, McKay VR*
*Public Health Information Seeking, Trust, and COVID-19 Prevention Behaviors: Cross-sectional Study*
*J Med Internet Res 2022;24(9):e37846*
*URL: https://www.jmir.org/2022/9/e37846*
*doi:10.2196/37846*
*PMID:36084197*

Review

# Chatbots for Smoking Cessation: Scoping Review

Robyn Whittaker[1,2], MBChB, MPH, PhD; Rosie Dobson[1], BSc, MSc, PGDipHlthPsych, PhD; Katie Garner[1], BSc

[1]National Institute for Health Innovation, University of Auckland, Auckland, New Zealand
[2]Waitemata District Health Board, Auckland, New Zealand

**Corresponding Author:**
Robyn Whittaker, MBChB, MPH, PhD
National Institute for Health Innovation
University of Auckland
22-30 Park Ave
Grafton
Auckland, 1142
New Zealand
Phone: 64 9373 7599
Email: r.whittaker@auckland.ac.nz

## Abstract

**Background:** Despite significant progress in reducing tobacco use over the past 2 decades, tobacco still kills over 8 million people every year. Digital interventions, such as text messaging, have been found to help people quit smoking. Chatbots, or conversational agents, are new digital tools that mimic instantaneous human conversation and therefore could extend the effectiveness of text messaging.

**Objective:** This scoping review aims to assess the extent of research in the chatbot literature for smoking cessation and provide recommendations for future research in this area.

**Methods:** Relevant studies were identified through searches conducted in Embase, MEDLINE, APA PsycINFO, Google Scholar, and Scopus, as well as additional searches on JMIR, Cochrane Library, Lancet Digital Health, and Digital Medicine. Studies were considered if they were conducted with tobacco smokers, were conducted between 2000 and 2021, were available in English, and included a chatbot intervention.

**Results:** Of 323 studies identified, 10 studies were included in the review (3 framework articles, 1 study protocol, 2 pilot studies, 2 trials, and 2 randomized controlled trials). Most studies noted some benefits related to smoking cessation and participant engagement; however, outcome measures varied considerably. The quality of the studies overall was low, with methodological issues and low follow-up rates.

**Conclusions:** More research is needed to make a firm conclusion about the efficacy of chatbots for smoking cessation. Researchers need to provide more in-depth descriptions of chatbot functionality, mode of delivery, and theoretical underpinnings. Consistency in language and terminology would also assist in reviews of what approaches work across the field.

***(J Med Internet Res 2022;24(9):e35556)*** doi:10.2196/35556

**KEYWORDS**

chatbot; conversational agent; COVID-19; smoking cessation

## Introduction

### Background

Tobacco use is the single most preventable cause of premature deaths worldwide. Tobacco use kills over 8 million people every year and is a risk factor for cancer, respiratory disease, cerebral vascular disease, heart disease, and a number of other debilitating chronic diseases [1]. Quitting smoking can lower the risk of some tobacco-related illnesses, and in some cases, ex-tobacco users can gain years in life expectancy compared to those who continue to use tobacco. Despite significant progress over the past 2 decades, with global tobacco use falling from 1.397 billion in 2000 to 1.337 billion in 2018 or approximately 60 million people [2], many countries are still not adequately implementing policies that can save lives from tobacco use, and the global target set by governments to reduce the prevalence of tobacco use by 30% by 2025 remains off track. One of those policies is World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) Article 14, which offers help to quit tobacco.

COVID-19 is an infectious disease that primarily affects the lungs, and tobacco smoking is a known risk factor for many respiratory infections and increases the severity of respiratory diseases. Early in the pandemic, the WHO convened a group of experts and found that severe COVID-19 is more likely to develop in smokers than nonsmokers. This news triggered millions of tobacco users to want to quit [3]. Despite approximately 60% of tobacco smokers voicing the desire to quit, only 30% have access to resources that will enable them to quit successfully [3].

Harnessing the power of digital solutions may be one approach to bridging this gap in the availability and delivery of cessation services globally. The subscription statistics of global telecommunications companies show that there are currently approximately 3 billion smartphone users, a number that is expected to grow by millions in the next few years [4]. This rapid growth in smartphone ownership and internet access has paved the way for new opportunities to reach tobacco smokers and provide smoking cessation support. Digital interventions have significant advantages for health care, such as flexibility, anonymity, cost-effectiveness, scalability, and increased access [5,6].

Previous research has found that SMS text messaging interventions are effective for helping tobacco users to quit successfully. Systematic reviews of high-quality (low-bias) randomized controlled trials (RCTs) of smartphone smoking cessation programs have found that SMS text messaging interventions generally double the success rate of abstinence [7,8]. A 2019 Cochrane review found beneficial results for 6-month cessation outcomes with text messaging, while there was insufficient evidence at that point for smartphone app interventions [8].

Most SMS programs use behavior change techniques (BCTs), which have been found to contribute to intervention efficacy [9]. Techniques from the COM-B model, which describes that a person's capability, motivation, and opportunity all contribute to their behavior [10], such as "facilitate relapse prevention and coping," "provide information on the consequences of smoking and smoking cessation," "promote the use of relaxation techniques," and "advise on/facilitate the use of social support," can be easily integrated into messaging interventions [9].

These BCTs that have been proven effective could also be used in chatbots or conversational agents, which are digital tools that allow users to talk with the program through voice or text [11]. These tools are designed to mimic human communication and provide friendly and engaging answers to directly respond to the user's questions or concerns [12]. In recent times, chatbots have been used in health care for supporting clinicians in diagnosis and helping individuals manage chronic illnesses [13]. Because chatbots allow instantaneous feedback and are widely available through social media and messaging apps, such as Facebook Messenger, it is hypothesized that they may be particularly suited to support smoking cessation, building on previous text messaging research.

During the COVID-19 pandemic, the WHO developed several smoking cessation chatbots, using content from existing smoking cessation programs. These were launched in response to the WHO initiative to get 100 million people to quit smoking, and they include Florence (a virtual human) and chatbots delivered on Viber, WhatsApp, Facebook Messenger, and WeChat, although formal evaluations of these are not yet available.

### Current Review

Despite the current potential of chatbots to help tobacco users quit, there are no reviews to assess their efficacy among adult smokers. The purpose of this scoping review is to assess the efficacy and summarize the research on chatbots for smoking cessation. The review includes a search of published and grey literature to examine the full extent of evidence.

## Methods

### Identification of Studies

This review was guided by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist [14], and the protocol was not registered. Embase, MEDLINE, APA PsycINFO, Google Scholar, and Scopus were searched on July 2, 2021. We also ran an additional search on JMIR, Cochrane Library, Lancet Digital Health, and Digital Medicine to identify any more relevant articles. An iterative process was used to develop the search strategy. Words associated with smoking cessation were combined with words for chatbots or "conversational agents" (see Textbox 1 for all search terms for each database). These combinations were used to create search strings, which were then put into the databases. Filters were used to limit the search to articles published between 2000 and 2021. The reference lists of included articles were also searched for any new studies that met the inclusion criteria.

**Textbox 1.** Search terms.

**Embase**

1. smoking cessation.mp. or exp Smoking Cessation/

2. (((quit$ or stop$ or ceas$ or giv$ or prevent$) adj3 smok$) or cigarette$).ti,ab.

3. exp passive smoking/

4. exp smoking habit/

5. smokeless tobacco/

6. smoking reduction/

7. (smok* or tobacco).mp.

8. 1 or 2 or 3 or 4 or 5 or 6 or 7

9. (Conversational agent$ or embodied conversational agent or chatbot$ or avatar or dialog$ system or virtual assistan$ or virtual nurs$ or virtual patient or virtual coach$ or intelligent assistan$ or relation$ agent or assistance technol$ or voice-based interfac$ or virtual coach or speech recognition software or voice recognition software).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]

10. 8 and 9

**MEDLINE**

1. smoking cessation.mp. or exp Smoking Cessation/

2. (((quit$ or stop$ or ceas$ or giv$ or prevent$) adj3 smok$) or cigarette$).ti,ab.

3. exp passive smoking/

4. exp smoking habit/

5. smokeless tobacco/

6. smoking reduction/

7. (smok* or tobacco).mp.

8. 1 or 2 or 3 or 4 or 5 or 6 or 7

9. (Conversational agent$ or embodied conversational agent or chatbot$ or avatar or dialog$ system or virtual assistan$ or virtual nurs$ or virtual patient or virtual coach$ or intelligent assistan$ or relation$ agent or assistance technol$ or voice-based interfac$ or virtual coach or speech recognition software or voice recognition software).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

10. 8 and 9

**APA psychINFO**

1. smoking cessation.mp. or exp Smoking Cessation/

2. (((quit$ or stop$ or ceas$ or giv$ or prevent$) adj3 smok$) or cigarette$).ti,ab.

3. exp passive smoking/

4. smokeless tobacco/

5. smoking reduction/

6. (smok* or tobacco).mp.

7. 1 or 2 or 3 or 4 or 5 or 6

8. (Conversational agent$ or embodied conversational agent or chatbot$ or avatar or dialog$ system or virtual assistan$ or virtual nurs$ or virtual patient or virtual coach$ or intelligent assistan$ or relation$ agent or assistance technol$ or voice-based interfac$ or virtual coach or speech recognition software or voice recognition software).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures, mesh]

9. 7 and 8

## Eligibility

Articles included in this scoping review met the following requirements: (1) conducted in tobacco smokers, (2) conducted between 2000 and 2021, (3) available in English, and (4) included a chatbot intervention. As this was a scoping review, we did not put any restrictions on study design; however, thesis manuscripts were excluded.

## Study Selection

Once duplicates were removed using EndNote, the titles were screened according to the inclusion criteria by 2 authors (KG and RD). The software used for this was Rayyan [15], a web program for systematic and scoping reviews. During this screening process, titles were screened to assess their relevance to the research question and criteria. If the relevance could not be inferred from the titles, the abstracts of the selected studies were reviewed according to the inclusion criteria. The next step was to assess the full text of articles that met the inclusion criteria and assess their relevance. Finally, the 2 researchers (KG and RD) met to discuss any disagreements and resolve any discrepancies. If necessary, a third researcher was consulted.

### Data Extraction and Synthesis

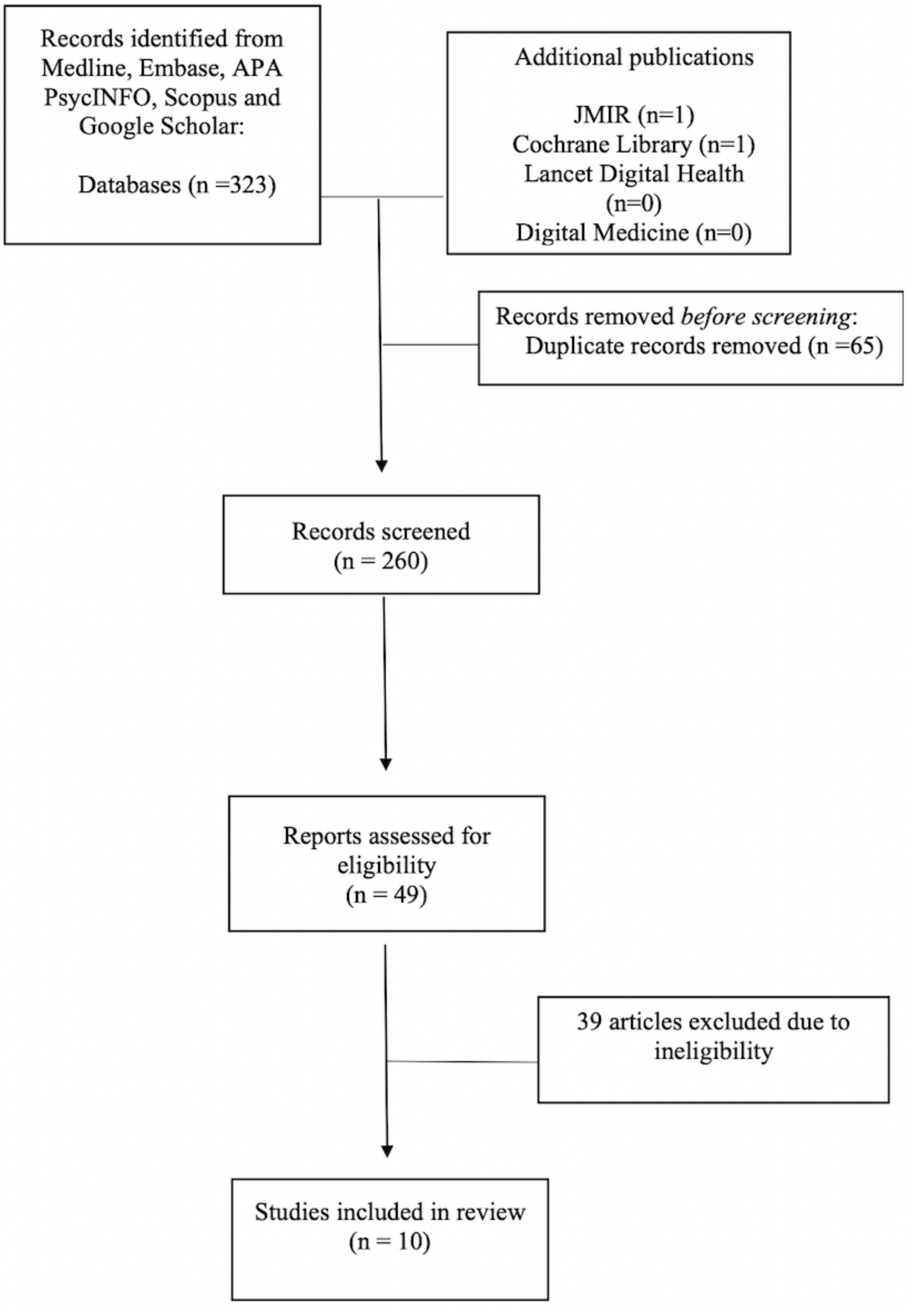
Two authors (KG and RD) extracted the data from the included studies on July 14, 2021. A predesigned spreadsheet was used to extract the data, including the title, author(s), total sample size, country, age mean and SD, intervention duration, theory referenced, measures, assessment timepoints, comparison condition, and main results, as well as whether a power calculation was conducted.

## Results

### Study Selection

The database search identified a total of 323 studies, and an additional 2 studies were found from Cochrane Library and JMIR (Figure 1). After duplicates were removed, 259 articles were screened. After the final review, 10 studies were included. Studies were excluded for several reasons: (1) the conversational agent was actually a human communicating via social media, (2) the paper was a thesis, and (3) the intervention did not target smoking cessation specifically.

**Figure 1.** Flowchart of the literature search and article selection. JMIR: *Journal of Medical Internet Research*.



### Study Characteristics

There was a mix of 3 framework articles [16-18], 1 study protocol [19], 2 pilot studies [20,21], 2 trials [22,23], and 2 RCTs [24,25]. The studies included in this review are presented in Table 1, and further details are provided in Multimedia Appendix 1.

Three papers were predominantly descriptive and did not present any interventions [16-18]. These described an array of topics in the smoking cessation chatbot literature, including single-agent and multiagent chatbots based on cognitive behavioral therapy [16]; a description of designing an embodied conversational agent for smoking cessation [17]; and research, theory, and considerations for smoking cessation chatbots in adolescents with a low socioeconomic status [18].

**Table 1.** Studies included in the review.

| Study | Type of paper | Chatbot intervention | Theory |
|---|---|---|---|
| Abdullah et al [22] | Study trial | Simple embodied conversational program that encourages smokers to think about setting a quit date. When participants log in, the agent initiates a discussion based on recorded information in the previous session. | N/A[a] |
| Almusharraf et al [23] | Study trial | Chatbot based on motivational interviewing. The chatbot delivers questions about the pros and cons of smoking. | Motivational interviewing |
| Avila-Tomas et al [19] | Study protocol | Chatbot that guides users through the stages of the dishabituation process. Includes cognitive-behavioral, relapse-prevention, and problem-solving techniques. | Gamification, cognitive behavioral theory, problem-solving |
| Calvaresi et al [16] | Framework paper | Describes a single-agent and multiagent chatbot based on CBT[b] and MAS[c] (underlying technology that models human-like behaviors). | CBT |
| Grolleman et al [17] | Framework paper | Describes designing an embodied conversational agent for smoking cessation. | Motivational interviewing and nonverbal listening techniques |
| Karekla et al [20] | Pilot randomized clinical trial | Digital avatar–led ACT[d] smoking cessation program. The program ran for 6 sessions, 25 minutes each. The avatar provides questions, and users answer. Involved ACT-based activities and homework exercises. | ACT |
| Masaki et al [21] | Single-arm pilot study | Smoking cessation app called CureApp that includes an artificial intelligence nurse to which users can send a message when they have a craving or are going through withdrawal. The nurse will respond with personalized advice and how to deal with the symptom. | N/A |
| Perski et al [24] | Randomized controlled trial | Smoking cessation chatbot that guides users through the UK “Stop Smoking Services” standard smoking cessation program. It checks in with users twice a day and is available for support when needed. Uses positive reinforcement for smoke-free days, resisting cravings, and quit milestones. | Behavior change techniques |
| Simon et al [18] | Descriptive summary paper | Chatbots for smoking cessation among adolescents with a low socioeconomic status. Describes interventions in motivational interviewing and CBT, as well as some gaps in the literature and future recommendations. | Motivational interviewing and CBT |
| Wang et al [25] | Randomized controlled trial | Chatbot on a WeChat support group. The chatbot sends announcements, reminders, ideas, and responses to the support group. | N/A |

[a] N/A: not applicable.

[b] CBT: cognitive behavioral therapy.

[c] MAS: multiagent systems.

[d] ACT: acceptance and commitment therapy.

### *Sample Size and Intervention Duration*

The pilot studies [20,21] and trials [22,23] had samples that ranged from 6 to 121 participants. With respect to the RCTs, the samples were reasonably large (n=57,214 [24] and n=401 [25]). The intervention duration for the RCTs was relatively short at only 1 to 2 months, and the longest follow-up period was 52 weeks in 1 pilot study [21].

### *Participant Characteristics*

Each study included participants who were adult tobacco smokers. These participants were recruited from countries across Europe, America, and Asia. There were no studies in low- and middle-income countries (LMICs).

### *Measures and Use of Theory*

Some of the studies used a validated measure, for example, the quality of life scale EuroQol-5D-5L (5-level EuroQol 5-dimensional questionnaire) [19] or the nicotine dependence scale (Fagerstrom Test for Nicotine Dependence) [20,21]. However, a number of studies used self-report measures to assess smoking as the postintervention measure [24,25]. Out of 9 studies, 7 referenced at least one theory. The theories included psychological theories such as motivational interviewing, acceptance and commitment therapy, cognitive behavioral therapy, BCTs, gamification, and problem-solving.

### *Types of Chatbots*

The chatbot interventions varied significantly across studies. Some chatbots were described as embodied conversational agents, which used an avatar in the form of a person to talk with the user [17,20,22]. Others used a social media platform, such as WhatsApp [25], or were part of an app on the mobile device [21,24]. The additional studies described the chatbot as a function on its own, which delivered smoking cessation support or delivered a smoking cessation program via a website or mobile device [16,19,23].

## Study Outcomes

### *Smoking Cessation*

A number of studies found positive outcomes for smoking cessation measures. Four studies noted improvements in quit rates. Karekla et al [20] found that the treatment group had significantly higher self-reported quit rates compared with controls at 6 months. Masaki et al [21] found that the chatbot intervention led to continuous abstinence rates of 64% from 9 to 24 weeks, 76% from 9 to 12 weeks, and 58% from 9 to 52 weeks. This was as a supplement to an outpatient nicotine dependence service, and these rates were said by the authors to be greater than those in previous studies of the outpatient service. Perski et al [24] also found that those who received a chatbot in support of their existing Smoke Free app had 1.41 times greater odds of being abstinent at 1 month than those who received the standard version of the app without the chatbot. Lastly, Wang et al [25] found that those in the intervention group had a higher quit rate compared with controls at 6 months (20.1% vs 12.4%). Although Almusharraf et al [23] did not measure smoking cessation, 8.3% of participants noted that the intervention helped with their smoking habits. Two studies found that the chatbot intervention led to a reduction in the number of cigarettes smoked per day after 14 days [22] and 6 months [20]. Other reported benefits to smoking cessation included lower levels of nicotine dependence at 6 months [20] and 12 weeks [21].

### *Additional Outcomes*

The studies in this scoping review also found some additional benefits worth highlighting. Three studies noted that the intervention was liked by participants. Abdullah et al [22] found that the overall appraisal was very positive, and Karekla et al [20] found that the satisfaction, interest, engagement, acceptability, and helpfulness of the intervention were rated highly. Almusharraf et al [23] found that 34.7% of participants enjoyed the interaction with the chatbot. In addition to these findings, Karekla et al [20] found that their chatbot intervention led to improvements in self-efficacy at 6 months compared with the control group. Masaki et al [21] found improvements in the mood and physical symptom scale at 12 weeks in the intervention group. Lastly, Wang et al [25] found improvements in engagement. When the chatbot was activated, researchers found an increase in the number and types of messages sent by 61%, and those who had not smoked in the past week had a significantly higher number of conversations with the chatbot compared with those who had smoked.

### *Study Quality*

A review of research methodology found that most studies were of low quality. Only 2 RCTs were found. In 1 RCT, only 10.7% of the overall sample responded at a 1-month follow-up [24]. The other RCT failed to report any P values [25]. In addition, the follow-up period was short (1 to 2 months). One framework paper highlighted some preliminary findings of a trial; however, the researchers did not provide many methodological details [16], and another trial had an attrition rate of 42% [20].

# Discussion

The aim of this scoping review was to provide a general summary of the extent of evidence for chatbots in smoking cessation. Overall, a summary of the evidence found that there are currently limited high-quality studies assessing chatbots for smoking cessation. Two RCTs provided promising results. However, both had significant methodological and reporting issues [24,25]. The trial and pilot studies also showed efficacy; however, due to the limited power of these studies, generalization to larger populations is limited. This scoping review suggests that larger-scale high-quality studies are needed to assess the effectiveness of chatbots for smoking cessation.

While the publication of reviews on the technical development and effectiveness of chatbots in general continues at pace [26-29], this scoping review highlights a number of key concerns specific to the evidence for the use of chatbots in smoking cessation. First, as discussed, the quality of evidence was low, with issues, such as short-term follow-up, high attrition/low response rates, and poorly described methodology. Another key concern was the absence of studies from LMICs. High-income countries are likely to have a different context, including good tobacco control policies and low smoking prevalence, making the generalizability of the findings of these studies to LMICs problematic. There were also no studies conducted in younger populations who may be more used to or willing to use chatbot technology. Simon et al [18] noted in their paper that smoking cessation interventions are needed for adolescents having a low socioeconomic status. Thus, future research will need to test these interventions in more diverse populations.

The last key concern was regarding the language used to describe chatbot interventions. The way researchers describe interventions varies significantly across studies. For example, some were described as “conversational agents,” “embodied conversational agents,” “digital avatars,” or apps that simply contained a chatbot. This heterogeneity makes it difficult to compare and accurately assess the efficacy of chatbot functionality. With no “agreed” terms and less than full descriptions of the functions these technologies provide, the likelihood of missing interventions that function like chatbots is high.

Further considerations raised by this scoping review include the continuum of chatbot technologies used in interventions. Some are relatively simple, for example, providing a set of responses/questions that participants choose from, which is not very different from existing messaging-based smoking cessation interventions that have been proven effective [30-32], while others are more complex and can respond with highly tailored responses to participants' individual questions using artificial intelligence methods. Chatbot technologies can be delivered in a variety of ways, from stand-alone apps and being embedded in existing social media/messaging programs to more sophisticated "digital human" technologies (eg, SoulMachines [33]). These variations may potentially impact engagement and effectiveness and thus should be fully described in any reports or studies.

It is also possible that chatbots can be used for other interventions related to, but not directly about, smoking cessation. We found a paper that described chatbots for screening substance users [34]. Chatbots may function well in helping with health-seeking behaviors and accessing health services before providing actual smoking cessation assistance.

While this scoping review provides an effective summary of the current chatbot literature, there are some limitations of this review. We did not conduct a formal assessment of the methodological quality of the literature [14] owing to the obvious lack of sufficient quality trials. There was also considerable heterogeneity across the chatbot interventions with respect to the description, level of contact with the intervention, methodology, and outcome measures. This made it difficult to assess the overall efficacy of chatbots and to synthesize the findings.

Despite these limitations, this review has some significant strengths. This review searched published and grey literature and did not place any restrictions on the types of studies included. As a result, we were able to assess a variety of studies, which provided a good summary of the current state of the research.

It is evident that more high-quality trials need to be performed to fully assess the efficacy of chatbots for smoking cessation. In addition, it is hoped that formal evaluations of smoking cessation chatbots developed by the WHO during the COVID-19 pandemic will provide real-world evidence to contribute to the body of knowledge about the impact of chatbots for smoking cessation.

In conclusion, this scoping review provides a summary of the use of chatbots for smoking cessation. Overall, we found some evidence for effectiveness; however, the number of studies assessing chatbots is limited, and it is evident that more high-quality trials are needed. Future research should aim to provide more in-depth descriptions of chatbot functionality, mode of delivery, and theoretical underpinnings. It should compare chatbots with proven text messaging and other cessation interventions to determine whether they can be more effective than current programs. Finally, authors should use consistent terminology in their descriptions of chatbots and in the keywords to ensure their studies are easily searchable for future reviews.

## Acknowledgments

The World Health Organization funded this work.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the included studies.
[DOCX File , 19 KB - jmir_v24i9e35556_app1.docx ]

## References

1. Tobacco. World Health Organization. URL: http://who.int/news-room/fact-sheets/detail/tobacco [accessed 2022-09-14]
2. WHO global report on trends in prevalence of tobacco use 2000-2025, fourth edition. World Health Organization. URL: https://www.who.int/publications/i/item/9789240039322 [accessed 2021-12-08]
3. WHO statement: Tobacco use and COVID-19. World Health Organization. 2020. URL: https://www.who.int/news/item/11-05-2020-who-statement-tobacco-use-and-covid-19 [accessed 2021-12-08]
4. Number of smartphone subscriptions worldwide from 2016 to 2021, with forecasts from 2022 to 2027. Statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2022-09-14]
5. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
6. Domhardt M, Steubl L, Baumeister H. Internet- and mobile-based interventions for mental and somatic conditions in children and adolescents. Z Kinder Jugendpsychiatr Psychother 2020 Jan;48(1):33-46 [FREE Full text] [doi: 10.1024/1422-4917/a000625] [Medline: 30422059]
7. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]

8. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
9. Michie S, Free C, West R. Characterising the 'Txt2Stop' smoking cessation text messaging intervention in terms of behaviour change techniques. J. Smok Cessat 2012 Aug 01;7(1):55-60. [doi: 10.1017/jsc.2012.12]
10. 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]
11. Lucas G, Gratch J, King A, Morency L. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 2014 Aug;37:94-100. [doi: 10.1016/j.chb.2014.04.043]
12. Brandtzaeg PB, Følstad A. Why People Use Chatbots. In: Internet Science. INSCI 2017. Lecture Notes in Computer Science, vol 10673. Cham: Springer; 2017:377-392.
13. Laranjo L, Dunn A, Tong H, Kocaballi A, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
14. Tricco A, Lillie E, Zarin W, O'Brien K, 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]
15. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210-210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
16. Calvaresi D, Calbimonte JP, Dubosson F, Najjar A, Schumacher M. Social Network Chatbots for Smoking Cessation: Agent and Multi-Agent Frameworks. In: WI '19: IEEE/WIC/ACM International Conference on Web Intelligence. 2019 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence; October 14-17, 2019; Thessaloniki, Greece p. 286-292. [doi: 10.1145/3350546.3352532]
17. Grolleman J, van Dijk B, Nijholt A, van Emst A. Break the Habit! Designing an e-Therapy Intervention Using a Virtual Coach in Aid of Smoking Cessation. In: IJsselsteijn WA, de Kort YAW, Midden C, Eggen B, van den Hoven E, editors. Persuasive Technology. PERSUASIVE 2006. Lecture Notes in Computer Science, vol 3962. Berlin, Heidelberg: Springer; 2006:133-141.
18. Simon P, Krishnan-Sarin S, Huang TH. On Using Chatbots to Promote Smoking Cessation Among Adolescents of Low Socioeconomic Status. arXiv. 2019. URL: https://arxiv.org/abs/1910.08814 [accessed 2022-09-14]
19. Ávila-Tomá J, Espinosa E, Lorenzo C, Suberviola F, Pardo B, Serrano M, Grupo Dej@lo. Dejal@Bot: Un chatbot aplicable en el tratamiento de la deshabituación tabáquica. Revista de Investigación y Educación en Ciencias de la Salud (RIECS) 2020;5(1):33-41 [FREE Full text]
20. Karekla M, Savvides S, Gloster A. An avatar-led intervention promotes smoking cessation in young adults: A pilot randomized clinical trial. Ann Behav Med 2020 Oct 01;54(10):747-760 [FREE Full text] [doi: 10.1093/abm/kaaa013] [Medline: 32383736]
21. Masaki K, Tateno H, Kameyama N, Morino E, Watanabe R, Sekine K, et al. Impact of a novel smartphone app (CureApp Smoking Cessation) on nicotine dependence: Prospective single-arm interventional pilot study. JMIR Mhealth Uhealth 2019 Feb 19;7(2):e12694 [FREE Full text] [doi: 10.2196/12694] [Medline: 30777848]
22. Abdullah A, Gaehde S, Bickmore T. A tablet based embodied conversational agent to promote smoking cessation among veterans: A feasibility study. J Epidemiol Glob Health 2018 Dec;8(3-4):225-230 [FREE Full text] [doi: 10.2991/j.jegh.2018.08.104] [Medline: 30864768]
23. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: Design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020 Nov 03;22(11):e20251 [FREE Full text] [doi: 10.2196/20251] [Medline: 33141095]
24. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digital Health 2019 Sep 30;5:205520761988067. [doi: 10.1177/2055207619880676]
25. Wang H, Zhang Q, Ip M, Fai Lau J. Social media–based conversational agents for health management and interventions. Computer 2018 Aug;51(8):26-33. [doi: 10.1109/MC.2018.3191249]
26. Rapp A, Curti L, Boldi A. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. International Journal of Human-Computer Studies 2021 Jul;151:102630. [doi: 10.1016/j.ijhcs.2021.102630]
27. Wang J, Hwang G, Chang C. Directions of the 100 most cited chatbot-related human behavior research: A review of academic publications. Computers and Education: Artificial Intelligence 2021;2:100023. [doi: 10.1016/j.caeai.2021.100023]
28. Borsci S, Malizia A, Schmettow M, van der Velde F, Tariverdiyeva G, Balaji D, et al. The Chatbot Usability Scale: the design and pilot of a usability scale for interaction with AI-based conversational agents. Pers Ubiquit Comput 2021 Jul 21;26(1):95-119. [doi: 10.1007/s00779-021-01582-9]
29. Oh Y, Zhang J, Fang M, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act 2021 Dec 11;18(1):160 [FREE Full text] [doi: 10.1186/s12966-021-01224-6] [Medline: 34895247]

30. Bramley D, Riddell T, Whittaker R, Corbett T, Lin R, Wills M, et al. Smoking cessation using mobile phone text messaging is as effective in Maori as non-Maori. N Z Med J 2005 Jun 03;118(1216):U1494. [Medline: 15937529]
31. Chen J, Ho E, Jiang Y, Whittaker R, Yang T, Bullen C. Mobile social network-based smoking cessation intervention for Chinese male smokers: Pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Oct 23;8(10):e17522 [FREE Full text] [doi: 10.2196/17522] [Medline: 33095184]
32. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin R, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
33. Overview of Our Digital Brain: How the Soul Machines Digital Brain is autonomously animating Digital People. Soul Machines. URL: https://www.soulmachines.com/2021/09/overview-of-our-digital-brain-how-the-soul-machines-digital-brain-is-autonomously-animating-digital-people/ [accessed 2021-12-09]
34. Auriacombe M, Moriceau S, Serre F, Denis C, Micoulaud-Franchi J, de Sevin E, et al. Development and validation of a virtual agent to screen tobacco and alcohol use disorders. Drug Alcohol Depend 2018 Dec 01;193:1-6 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.08.025] [Medline: 30321739]

## Abbreviations

**BCT:** behavior change technique
**LMIC:** low- and middle-income country
**RCT:** randomized controlled trial
**WHO:** World Health Organization

*Edited by T Leung, V Arnold, M Reinfelde; submitted 09.12.21; peer-reviewed by O Perski, K Cummins; comments to author 23.03.22; revised version received 12.04.22; accepted 18.08.22; published 26.09.22.*

*Please cite as:*
*Whittaker R, Dobson R, Garner K*
*Chatbots for Smoking Cessation: Scoping Review*
*J Med Internet Res 2022;24(9):e35556*
*URL: https://www.jmir.org/2022/9/e35556*
*doi:10.2196/35556*
*PMID:36095295*

Research Letter

# Digitalization of Health Care: Findings From Key Informant Interviews in Sweden on Technical, Regulatory, and Patient Safety Aspects

Björn Ekman[1], PhD; Hans Thulesius[1], PhD; Jens Wilkens[1], MA; Eva Arvidsson[2], PhD

[1]Department of Clinical Sciences, Lund University, Malmö, Sweden

[2]Futurum, Region Jönköpings län, Jönköping, Sweden

**Corresponding Author:**
Björn Ekman, PhD
Department of Clinical Sciences
Lund University
Jan Waldenströms Gata 35
Malmö, 20213
Sweden
Phone: 46 727145757
Email: bjorn.ekman@med.lu.se

***(J Med Internet Res 2022;24(9):e38746)*** doi:10.2196/38746

**KEYWORDS**

digitalization; regulation; patient safety; quality of care; interview; Sweden; health care; digital health; decision-making; digital care

## Introduction

Over the past few decades, digital technologies for health care have been introduced to improve effectiveness, reduce costs, and enhance access [1]. The case for digital health care increased dramatically during the COVID-19 pandemic [2,3]. However, less is known about some of the broader conditions under which these technologies would be able to contribute to improving health care [4]. To inform policy development from a national perspective, we conducted key informant interviews around 3 domains of digitalization: (1) infrastructure and skills, (2) regulatory considerations, and (3) patient safety and quality of care (see, for example, Desveaux et al [4]).

## Methods

### Overview

In spring 2021, we interviewed 24 representatives identified in a scoping exercise across 4 areas: providers of health services, researchers active in the field, regional health authorities, and national-level organizations. The responses were analyzed using conventional content analysis [5].

### Ethics Approval

The Swedish Ethical Review Authority reviewed the study (2020-04381; September 29, 2020).

## Results

### Infrastructure and Skills

The technical conditions for health care digitalization in Sweden are favorable. Broadband is generally available; most people have access to a computer, tablet, or smartphone; and the level of acceptability of using such tools is high. The existence of a national digital system for secure identification was also noted.

However, challenges and concerns were also noted, including variation in access and ability to navigate digital tools across population groups. Older adults, immigrants, and persons living with a disability may have difficulties accessing services by means of digital tools. The COVID-19 pandemic has accelerated the application of digital primary care in Sweden [6].

### Regulatory Aspects

Respondents noted the largely weak regulatory environment for health care in Sweden, including the absence of national policies for digital care and a host of regulators with unclear mandates and roles.

The lack of a strategic approach at the national level and the strong decentralization and autonomy of the 21 Swedish health care regions lead to significant local variation in the approach to digitalization in health care. Examples included a plethora of digital platforms for health services, different requirements with respect to public purchasing of services, and variations in the application of the need for data sharing and storage. A particular concern was the case-based payment for digital care while traditional primary care services are reimbursed through capitation.

### Patient Safety and Quality-of-Care Issues

Respondents noted both opportunities and risks with respect to patient safety and quality of care. On the one hand, digital technologies were seen as safe and may support improved quality of care as more data are available for clinical decision-making. On the other hand, there may be limitations to telemedicine for certain conditions and for patients with heavier demands, suggesting a further need to evaluate the effects. The risk of data mismanagement when using digital models of health care was also noted.

However, respondents also noted that there are similar risks with existing in-person models of care and that it is partly a question of finding the right balance between using digital technologies in a way that maximizes their effectiveness while managing the risks of such technologies. During the COVID-19 pandemic, digital care proved to be the safest option.

## Discussion

### Principal Findings

The continued integration of advanced digital technologies in health care is likely to transform much of medical service delivery, both in scope and content. Understanding the implications across the technical, regulatory, and patient safety aspects will be critical for the successful digitalization of health care.

The findings suggest that ensuring the technical conditions for digital care may be challenging even in a relatively advanced country such as Sweden. In resource-limited countries, this may be even more so [7]. The COVID-19 pandemic may have exposed some of these issues [6].

Failure to ensure a clear, effective, and transparent regulatory environment for digital care may risk jeopardizing the realization of the opportunities that digital technologies present and aggravating the risks that such technologies also entail. These factors relate critically to patient safety and to the quality of care. Ongoing analysis of the material using grounded theory will extend the analysis and deepen the understanding of these issues.

### Conclusion

The successful application of digital technologies for health care will depend on the strategic management of a set of technical, regulatory, and patient safety aspects. Even in advanced countries, there may be concerns related to each of these. Policy makers and regulatory agencies need to be made aware of these challenges. Further research will be needed to advance the understanding of the digitalization of health care along these dimensions.

---

### Conflicts of Interest

None declared.

### References

1. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021 Jan 26;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
2. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
3. Wong BLH, Maaß L, Vodden A, van Kessel R, Sorbello S, Buttigieg S, European Public Health Association (EUPHA) Digital Health Section. The dawn of digital public health in Europe: Implications for public health policy and practice. Lancet Reg Health Eur 2022 Mar;14:100316 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100316] [Medline: 35132399]
4. Desveaux L, Soobiah C, Bhatia RS, Shaw J. Identifying and overcoming policy-level barriers to the implementation of digital health innovation: qualitative study. J Med Internet Res 2019 Dec 20;21(12):e14994 [FREE Full text] [doi: 10.2196/14994] [Medline: 31859679]
5. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
6. Ekman B, Arvidsson E, Thulesius H, Wilkens J, Cronberg O. Impact of the Covid-19 pandemic on primary care utilization: evidence from Sweden using national register data. BMC Res Notes 2021 Nov 24;14(1):424 [FREE Full text] [doi: 10.1186/s13104-021-05839-7] [Medline: 34819161]
7. Opoku D, Busse R, Quentin W. Achieving sustainability and scale-up of mobile health noncommunicable disease interventions in Sub-Saharan Africa: views of policy makers in Ghana. JMIR Mhealth Uhealth 2019 May 03;7(5):e11497 [FREE Full text] [doi: 10.2196/11497] [Medline: 31066706]

*Edited by T Leung; submitted 15.04.22; peer-reviewed by L Maaß, J Petersen, T Prætorius; comments to author 01.06.22; revised version received 15.06.22; accepted 13.07.22; published 01.09.22.*

*Please cite as:*
*Ekman B, Thulesius H, Wilkens J, Arvidsson E*
*Digitalization of Health Care: Findings From Key Informant Interviews in Sweden on Technical, Regulatory, and Patient Safety Aspects*
*J Med Internet Res 2022;24(9):e38746*
*URL: https://www.jmir.org/2022/9/e38746*
*doi:10.2196/38746*
*PMID:36048503*

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org